Molecular biomarkers for seabird age estimation

Implications for ecological monitoring

Ricardo De Paoli-Iseppi

BSc (Hons), BComp, University of Tasmania

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

University of Tasmania

January 2019

Thesis abstract

Seabirds are widely used as an indicator species to track important changes in ecosystems. The key data required for monitoring include (i) population trends (ii) status, including demographic properties such as age structure and reproductive performance. Most seabirds have no outward or easily identifiable marks to age individuals and there are currently no robust molecular methods for age estimation in birds. Instead, individuals must be marked by leg rings as chicks to establish known-age populations, which is a labour intensive and expensive process. Long-term monitoring of some populations of birds has produced only small numbers of individuals for which there is a known age, highlighting the urgent need to identify alternative aging techniques.

In this thesis, I describe the development and testing of age-related differentially methylated positions (aDMPs) as a minimally invasive method for the estimation of chronological age in the Short-tailed shearwater (*Ardenna tenuirostris*). Recent studies have revealed novel examples of DNA methylation (DNAm) age association in several mammalian species, indicating the potential for developing DNAm age biomarkers for a broad range of wild animals. Emerging technologies for measuring epigenetic signals have also enhanced our ability to study age-related DNAm changes.

I assessed DNAm in several shearwater genes that have shown age-related DNAm changes in mammals. In birds ranging in age from chicks (months old) to 21 years, bisulphite treated blood and feather DNA was sequenced. The sequences allowed DNAm analysis across 67 CpG sites in 13 target gene regions. Despite the identification of some weakly correlated aDMPs, the majority had no clear association with age and statistical analysis using a penalised lasso approach did not produce an accurate ageing model. Our data indicated that some age-related signatures identified in orthologous mammalian genes are not conserved in the shearwater and that an alternative technique would be required to progress in this area.

The alternative approach that I tried was digital restriction enzyme analysis of methylation (DREAM). I quantified DNAm in whole blood samples from a total of 71 known-age shearwater using DREAM. This method measures DNAm levels at thousands of CpG dinucleotides throughout the genome. We identified seven CpG sites with DNAm levels that correlated with age. A model based on these relationships estimated age within a mean of 2.8

years of known age, based on validation estimates from models created by repeated sampling of training and validation data subsets. Longitudinal observation of individuals re-sampled over 1 or 2 years generally showed an increase in estimated age (6/7 cases). These seven aDMPs were then selected for use in a more cost-effective, targeted approach using a next generation sequencing assay. I validated and then generated chronological age estimates in a large set of known, minimum and unknown age individuals sourced from our study population.

Using targeted NGS we confirmed age-correlation in three of the aDMPs identified using the DREAM method. We also identified an additional 14 aDMPs that were within close proximity to two of the original markers described above. Using four aDMPs to create a shearwater epigenetic age assay we trained a model on 109 known-age blood samples. DNAm age estimates of minimum-age individuals (N = 55) showed that 39/55 (71%) had estimates above their known minimum-age. A further 23 unbanded individuals from a nearby island were used to create a small population age structure. This age structure, although small, showed similar age-class assignment to the known structure determined from long-term ringing.

For the first time, we have shown that epigenetic changes with age can be detected in a wild bird. The ability to differentiate age classes e.g. chicks, young breeders and birds of middle and old ages, of previously unbanded seabirds will be a useful tool for scientists and managers. Estimated age data could be informative for monitoring of post-pest eradication around island-breeding seabird populations or for determining impacts of longline fishery interactions. The molecular techniques developed for age estimation in this thesis also provide fundamental knowledge on seabird and mammal gene and CpG conservation, DNAm quantification and analysis in non-model animals.

Declaration of originality

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university and, to the best of my knowledge, contains no material previously published or written by any other person unless due reference is made in the text of this thesis.

Ricardo De Paoli-Iseppi Date: 29th January 2019

Statement of authority of access

The publishers of the papers comprising Chapters 2, 3 and 4 hold the copyright for that content and access to the material should be sought from the respective journals. Chapters 2 and 3 are open access and can be downloaded without restriction. The remaining non-published content of the thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

Ricardo De Paoli-Iseppi Date: 29th January 2019

Statement of ethical conduct

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University. The samples obtained for use in experimentation presented in this body of work were obtained following approval from the animal ethics committee (AEC) of the University of Tasmania, permits: A0016107 and A14277 and the Department of Primary Industries, Parks, Wildlife and Environment (DPIPWE) permits: FA18071, FA17340 and FA15230.

Ricardo De Paoli-Iseppi Date: 29th January 2019

Statement of Co-Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

- Ricardo De Paoli-Iseppi, University of Tasmania = **Candidate (RD)**
- Prof. Simon Jarman, CSIRO & Curtin University = **Author 1 (SJ)**
- Prof. Mark Hindell, Institute for Marine and Antarctic Studies = **Author 2 (MH)**
- Assoc. Prof. Joanne Dickinson, Menzies Institute for Medical Research = **Author 3 (JD)**
- Dr. Bruce Deagle, Australian Antarctic Division = **Author 4 (BD)**
- Dr. Clive McMahon, Sydney Institute for Marine Science = **Author 5 (CM)**
- Andrea Polanowski, Australian Antarctic Division = **Author 6 (AP)**

Author details and roles

- **Manuscript 1 (Chapter 2): De Paoli-Iseppi Ricardo**, Deagle Bruce E., McMahon Clive R., Hindell Mark A., Dickinson Joanne L., Jarman Simon N. (2017) Measuring Animal Age with DNA Methylation: From Humans to Wild Animals. Frontiers in Genetics. Vol. 8.
    - **RD 72.5%**, SJ 7.5%, MH 5%, JD 5%, CM 5%, BD 5%.
    - Conception: RD, JD, SJ; Design: RD, SJ, BD, MH, CM; Drafting original: RD, SJ, BD; Critical revision: JD, MH, CM; Final approval and accountability, all authors.

- **Manuscript 2 (Chapter 3): De Paoli-Iseppi R**, Polanowski AM, McMahon C, Deagle BE, Dickinson JL, Hindell MA, et al. (2017) DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. PLoS ONE 12(12): e0189181. https://doi.org/10.1371/journal.pone.0189181
    - **RD 70%**, SJ 7%, MH 5%, JD 5%, CM 5%, AP 3%, BD 5%.
    - Conception: RD, JD, SJ; Formal analysis: RD, BD, SJ; Funding acquisition: RD, MH, SJ, BD; Investigation: RD, AP; Methodology: RD, AP, SJ;

Software: RD, BD, SJ; Writing and critical revision: all authors.

- **Manuscript 3 (Chapter 4): De Paoli-Iseppi, R.**, Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A. and Jarman, S. N. (2018), Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. Molecular Ecology Resources. Accepted Author Manuscript. doi:10.1111/1755-0998.12981
    - **RD 72%**, SJ 5%, MH 5%, JD 5%, CM 5%, AP 3%, BD 5%.
    - Conception: RD, SJ; Formal analysis: RD, BD, SJ; Funding acquisition: RD, MH, SJ, BD; Investigation: RD, AP; Methodology: RD, AP; Software: RD, BD, SJ; Writing and critical revision: all authors.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Ricardo De Paoli-Iseppi, Candidate
Institute for Marine and Antarctic Studies, University of Tasmania
Date: 29th January 2019

Professor Simon Jarman, Supervisor
The University of Western Australia
Date: 29th January 2019

Professor Mark Hindell, Primary Supervisor
Institute for Marine and Antarctic Studies, University of Tasmania
Date: 29th January 2019

Professor Craig Johnson, Head of Ecology & Biodiversity Centre
Institute for Marine and Antarctic Studies, University of Tasmania
Date: 27th October 2019

Acknowledgements

This PhD has combined two of my scientific passions, epigenetics and zoology. It has been an interesting and enjoyable adventure with all the ups and downs that come with research. First, my excellent supervisors Simon Jarman, Joanne Dickinson, Clive McMahon and Mark Hindell; your combined breadth of knowledge and support throughout my PhD has been amazing, and I thank you for trying out such a cross-disciplinary project. Simon, thanks for creating this project, it has been rewarding and provided me with a lot of opportunities. Your guidance (near or far) has been fantastic, my scientific writing has improved out of sight and thanks for letting your student win at squash! Jo, thanks for three more years of your excellent advice and morning coffee chats, they're invaluable. Clive and Mark, thank you for helping me get my head around shearwater ecology, always being up for either a beer or sticking your arm down burrow holes, it was never a dull moment in the field.

This project, particularly the fieldwork, would not have been possible without the volunteer efforts of several fellow students, staff at IMAS, UTAS, AAD, DPIPWE and Parks and Wildlife Tasmania: Cassandra Price, Fernando Arce Gonzalez, Ross Monash, Robbie Gaffney, Eric Schwarz, Angela McGuire, Fiona Taylor and Parks field volunteers. I would also like to thank Richard Phillips and his team at the British Antarctic Survey and the genomic team at CIBIO for hosting me in 2017.

Funding for this project was provided by the Australian Antarctic Science grant (Project 4014), the New South Wales Linnean Society (Joyce W. Vickery Research Fund) and Holsworth Wildlife Research Endowment. Thanks to the Friends of Fisher Island for the loan of materials and transportation to Fisher Island.

Special mention to my research advisor Bruce Deagle for always being available to answer questions and have meandering discussions and lunchtime swims, your insight was extremely valuable. Andrea Polanowski, you are a powerhouse of the lab, thanks for teaching me so many new things and for your never-ending patience. James Marthick, you're a good mate and keep working on those dad jokes! Your expert supervision on the MiSeq ensured we didn't have a failed run. Julie McInnes, Leonie Suter and Laurence Clarke, cups of coffee, homemade cakes and banter was everything I needed to get going in the mornings and I wish you all the best.

Finally, thanks to my family and friends, without you and your support I wouldn't be sitting here finishing my thesis. Liam, Stephen and Sean, whether we were chasing a squash, cricket or soccer ball or eating burgers watching people do those sports better than we could, you always made me laugh and I couldn't ask for better company. Of course mum and dad for their love and support from the start, providing food, help with the house, travel, growing veggies and building things, it wouldn't have happened without the greatest parents going around. Leah, thanks for all your support, being my travel and dive buddy, wine provider and for flying backwards and forwards all this time, I'm looking forward to the next adventure with you.

Table of Contents

List of Figures

Chapter 5

List of Tables

Abbreviations

27k BeadChip Illumina	
450k BeadChip Illumina	
5mC	Methylated cytosine
aDMP	Age-related differentially methylated position
aDMR	Age-related differentially methylated region
bscDNA	Bisulphite converted DNA
CpG	Refers to a cytosine – guanine (CG) base pair
CpGi	CpG island
DNA	Deoxyribonucleic acid
DNAm	DNA methylation
DNMT	DNA methyltransferase
DREAM	Digital restriction enzyme analysis of methylation
gDNA	Genomic DNA
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HTS	High throughput sequencing
MAB	Molecular age biomarker
miRNA	Micro RNA
mRNA	Messenger RNA
mRS-PCR	Multiplexed restriction site PCR
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
RNA	Ribonucelic acid
RNAi	RNA interference
sjTREC	Signal joint T-cell recombination excision circles
UCSC	University of California Santa Cruz

Chapter 1 - Introduction

"Foaming and frothing from mountainous height,
Roaring like thunder the Rhyadr falls;
Though its silvery splendour the eye may delight,
Its fury the heart of the bravest appals."

George Borrow

Wild Wales

1.1 Chronological ageing

Chronological age is an important factor in animal ecology because many biological characteristics change with time. The effect of increasing chronological age on the biological age of an animal determines key ecological characteristics such as age of reproductive maturity (Charpentier *et al.* 2008; Essington *et al.* 2001; Nussey *et al.* 2008), reproductive frequency (Bradley & Safran 2014), likelihood of future reproduction (Jones *et al.* 2014; Massot *et al.* 2011) and mortality (Bradley & Safran 2014; Pérez-Barbería *et al.* 2014; Sherley *et al.* 2014). Examples of more specific age-dependent characteristics include the potential for pathogen transmission by mosquitoes to mammals (Cook *et al.* 2006; Wang *et al.* 2013), or growth and food consumption rates in fish (Essington *et al.* 2001). Age information is therefore indispensable for developing a full understanding of the ecology and life cycle of an animal species.

For animal populations, age-class distribution is a significant determinant of population growth rate (Tkadlec & Zejda 1998). Population age structures also reflect past population growth rates, for example an over-representation of young individuals in fast-growing populations; or of older individuals in shrinking populations (Ozgul *et al.* 2010). Population age structure can also be significantly altered by size-selective removal of individuals by natural predation or human exploitation and therefore reflect past demographic events (Campana 2001; Hilborn & Walters 1992; Polanowski *et al.* 2014). The proportion of a population that has reached the age of reproductive maturity often varies among populations (Charpentier *et al.* 2008). This in turn influences the ages at which first reproduction can occur and is an import predictor of population growth potential (Acker *et al.* 2014; Langvatn *et al.* 1996). Changes in the age of first reproduction can be a useful indicator of population status and pressures such as overfishing of long-lived species (Jennings *et al.* 1998). Age class distribution is often used in management of wild populations such as in setting catch quotas for fishery species (Hilborn & Walters 1992; Jennings *et al.* 1998) or in population viability analyses for endangered species (Beissinger & Westphal 1998).

1.1.2 Natural and artificial age markers

The importance of chronological age in animal ecology has driven the development of many methods for its estimation in wild animals. Morphological features of internal structures such as the annual growth rings in fish otoliths (Campana 2001) and mammalian teeth (Pérez‐Barbería *et al.* 2014); or daily growth rings in squid statoliths (Lipinski 1986) can provide

accurate age estimation from dead animals. Estimation of age in live animals is generally more challenging. Some species have external features that change predictably with age, such as the overall length of fish species with indeterminate growth (Essington *et al.* 2001) or the clear developmental stages found in many arthropod species that allow their sub-adult age to be determined (Amendt *et al.* 2011). However, many species have no external features that enable robust age estimation. Age can be established in some cases with a mark-recapture approach, whereby individuals are identified at birth so that their age is known when they are subsequently encountered (Clutton-Brock & Sheldon 2010). Individuals can also be recognised from artificial marks such as unique numbered bands attached to the legs of birds (Sherley *et al.* 2014), or natural markings like the pigmentation patterns on the ventral side of humpback whale flukes (Katona & Whitehead 1981). These longitudinal studies are only possible under certain circumstances and are labour intensive and costly, so estimating age by proxy markers has also been tested in a range of animal species.

1.1.3 Ageing and molecular change

Many aspects of biological ageing are under genetic control or result from genetic damage and are apparent in changes to abundance, sequence or epigenetic modifications to specific DNA or RNA regions (De Magalhaes & Costa 2009; Jones *et al.* 2015). Identifying these molecular changes in an animal species of interest and developing an assay for them is the primary challenge for molecular ecologists in developing MABs. Biological ageing follows multiple stages with diverse drivers and phenotypes at the organismal, cellular and molecular levels (Rando & Chang 2012).

Chemical damage to DNA appears to be a common characteristic for the life of most organisms and the source of this damage can be highly variable in nature, ranging from high levels of UV radiation, mitochondrial reactive oxygen species (ROS), pollution and other natural and artificial environmental factors (Egan & Zierath 2013). ROS are still considered to play a major role in ageing, however it does not account for all age-related phenotypes (Egan *et al.* 2010). Damage can interfere with critical cellular processes and whilst short-term responses may protect cells and tissues, constant exposure to the source of damage or accumulation of damage can lead to tissue or organ dysfunction. With some exceptions, long-lived species generally repair or prevent damage to critical areas with high efficiency so understanding how damage and ageing phenotypes present is difficult. Animal models of accelerated damage can display the early appearance of age phenotypes, and are a useful

method for exploring this relationship. There are exceptions to these models, perhaps indicating compensatory or adaptive mechanisms used to reduce or repair damage (Rönn *et al.* 2013). It is important to note that ageing is variable and can differ among individuals, even when environmental and genetic aspects are tightly controlled (Egan & Zierath 2013).

Development is the best understood aspect of the genetically controlled components of ageing. Studies of model organisms, such as *Drosophila melanogaster, Mus musculus* and *Caenorhabditis elegans*, have established conserved functions for many genetic pathways in animal development (Sempowski *et al.* 2002). Development either stops at maturity in animals with determinate growth or in animals with indeterminate growth, some aspects of development continue throughout the lifespan (Essington *et al.* 2001; Tissenbaum 2012). However, in most species developmental genes are primarily active during morphogenesis (Sempowski *et al.* 2002). Expression of these genes can be used as a MAB for identifying the developmental stage of an animal. For example, gene expression can be used for the identification of mosquitos that have reached the stage where they are able to transmit parasites to the animals from which they extract blood meals (Cook *et al.* 2006).

Nucleic acid changes that have been used as MABs across entire lifespans can be divided into two categories: accumulation of change in specific DNA molecules; and epigenetic changes that can alter the regulation of gene expression with age. Accumulated molecular damage is thought to be the primary driver of 'senescence', which in its most general definition is a loss of biological function with age. 'Cellular senescence' can be defined as the cessation of the cell cycle so that senescent cells remain in the G_1 phase, unable to replicate (Benhamed *et al.* 2012; Oliveira *et al.* 2014). Genes involved in regulating cellular senescence are good targets for MAB development.

1.2 Molecular age biomarkers (MABs)

Molecular age biomarkers (MABs) measure aspects of DNA or RNA sequence or abundance that change over the lifespan of an animal (Jarman *et al.* 2015). Molecular ecologists have focused so far on a small number of MAB types, most prominently telomere length assays, but for chronological age estimation these have only proven useful in cross-sectional studies of a limited number of species (Dunshea *et al.* 2011). Research on genetic aspects of the ageing process has accelerated in recent years as a result of technological advances allowing a range of new molecules to be characterised and quantified. Methods that improve the ability to study epigenetic changes associated with age have been particularly promising.

An ideal MAB for ecological applications would be measureable with exact precision and have a perfectly accurate relationship with chronological age for all individuals in a population. MABs are a proxy marker for chronological age and in reality all of them are influenced to some extent by other factors. Exogenous influences such as prevailing temperature (Wang *et al.* 2013) or food availability alter the expression of some developmental genes (Essington *et al.* 2001); as well as some genetic markers in fully developed animals (Tkadlec & Zejda 1998). Endogenous factors such as inherited MAB variation will also contribute to variance in a MABs relationship with age.

1.2.1 MABs and epigenetic regulation

Epigenetic mechanisms that regulate gene expression direct several processes involved in organismal and cellular ageing. The process of cell differentiation and 'canalization' into a fully differentiated cell type is largely controlled by epigenetic mechanisms (Jones *et al.* 2015). Epigenetic control ensures stable expression of genes that results in sets of cell-specific molecules for cell type differentiation. Recent research has also identified age-related epigenetic regulation of specific mammalian genes that are not part of developmental pathways, a phenomenon often referred to as the 'epigenetic clock' (Horvath 2013). Regulated epigenetic changes are distinct from genome-wide demethylation, which is produced by imperfect maintenance of the DNA methylation state through repeated mitoses. This phenomenon is called 'epigenetic drift' and is thought to act essentially randomly throughout the genome (Jones *et al.* 2015). Several genes that display epigenetic 'clock-type' regulation appear to have consistently altered DNA methylation throughout the animals' lifespan and some have been applied as MABs.

1.2.2 MAB development

Identification of a tissue-type that is practical to sample from the target species is a primary consideration in MAB development. For example, one of the most precise MABs so far developed involves qPCR of sjTREC levels in human blood (Ou *et al.* 2012). Applying this methodology to other species is appealing, but this MAB will only be useful in studies where suitable blood samples can be collected from the species of interest. The choice of which MAB to develop may also depend upon existing sample archives and preservation issues for collection of new samples. In many cases, genetic samples from the target species are already

available and it is likely that DNA-based MABs such as sjTRECs, DNA methylation or mitochondrial mutation accumulation could be directly applied to these samples.

RNA is less chemically stable than DNA, so RNA-based MABs such as mRNA and miRNA may be less appropriate for archived samples. RNA preservation technologies have greatly improved in recent years and with appropriate sample handling either DNA or RNA MABs could be applied in most studies where new samples are collected for MAB measurement (Gayral *et al.* 2011). The costs of MAB development and application may be another consideration when exploring age estimation possibilities (Jarman *et al.* 2015).

1.2.3 MAB calibration

Assessment of the relationship between chronological age and the response of a potential MAB requires tissue samples from animals of known age. Short-lived species can be kept in captivity for their entire lifespan to provide these samples, for example in insects (Cook *et al.* 2006; Wang *et al.* 2013), fish (Petzold *et al.* 2013), and small mammals (Maegawa *et al.* 2010). Suitable samples from long-lived species may come from zoos or aquaria, or the subjects of extensive longitudinal field studies. For example, many bird species have populations containing individuals of known age that were marked as chicks with leg bands (Sherley *et al.* 2014); species such as humpback whales have natural markings allowing them to be tracked throughout their lifespan (Katona & Whitehead 1981). Samples taken from known-age individuals in these populations can be used to calibrate a MAB that can be applied to any wild member of the same species (Polanowski *et al.* 2014).

A challenge for studies of long-lived animals is that longitudinal studies can have limited calibration data for the oldest ages, but this is not essential for all applications of age information in ecology (Campana 2001). All MABs developed to date have been calibrated from cross-sectional sampling of individuals with a range of ages in the calibration population. Resampling individuals at different times to measure change in a MAB is certainly possible, as has been demonstrated for mark-recapture calibration of morphological age estimators and this could be a useful MAB calibration approach for species where no known-age individuals exist (Bravington *et al.* 2014; Eaton & Link 2011).

If a single MAB has low accuracy it may be combined with other MABs, or non-MAB covariates with a method such as multiple linear regression to produce a combined assay based on multiple markers with an improved accuracy (Pauli *et al.* 2011). This approach is

commonly applied for MABs based on DNA methylation at multiple CpG sites (Bocklandt *et al.* 2011; Polanowski *et al.* 2014). However, low accuracy is an indicator that other co-factors such as environmental variables may be influencing the marker.

1.3 Epigenetic control of gene expression

1.3.1 Epigenetic modifications

Epigenetics is defined as modifications made to a DNA sequence that do not involve direct changes to the sequence itself and can be transmissible to daughter cells (Berger *et al.* 2009; Bird 2007). Chromosomes reside within the cell nucleus and are regulated assemblies of DNA and protein called chromatin. DNA wraps around histone proteins to form nucleosomes and these individual components can be epigenetically modified affecting DNA interactions and the binding of regulatory factors (Kouzarides 2007; Leung *et al.* 2012). Cells in eukaryotic organisms are genetically homogenous and epigenetic marks set up the framework for the correct development and differentiation of these cells. Given the dynamic nature of epigenetic modifications, epigenetics is commonly referred to as the interface between the genome and the environment (Feil & Fraga 2012). The packaging of DNA into nucleosomes already provides a layer of repression due to reduced DNA access, additional changes may then further restrict DNA binding access or reduce repression (Mohn & Schübeler 2009).

The self-renewal and differentiation processes of stem and progenitor cells requires the maintenance of a tightly controlled gene expression system which must be adjusted to a new fate upon differentiation (Mohn & Schübeler 2009). Epigenetic alterations and subsequent control are a critical component of these processes (Issa 2000). In mammals, post-synthetic modifications of DNA allow the genome flexibility to respond to both developmental and environmental cues. Histone acetylation is an epigenetic modification of lysine residues on the nucleosome core and is catalysed by histone acetyltransferase (HAT) (Bernstein *et al.* 2007). Acetylation of histones typically leads to a relaxed chromatin state, known as euchromatin allowing increased gene transcription. The removal of these acetyl groups by histone deacetylase (HDAC) leads to the opposite chromatin state known as heterochromatin, characterised by condensation of the DNA and reduced transcriptional activity. DNA methylation, an accessible and quantifiable epigenetic marker is commonly associated with reduced gene expression, the stability of gene expression states and for these reasons it is one of the most studied epigenetic marks.

Epigenetic control in humans is demonstrated in memory T cells which have the capacity for self-renewal and potential for effector functionality when challenged (Youngblood *et al.* 2013). During acute viral infection, antigen-specific T cells undergo large phenotype and functional changes, these changes in phenotype that accompany differentiation are in part controlled by epigenetic changes to DNA and histones (Youngblood *et al.* 2013). Early studies in these cells demonstrated that DNA methylation of *lfng* correlated with alterations in DNA structure and cytokine transcription (Farrar *et al.* 1985). Studies using HDAC inhibitors to identify epigenetic control over transcription for T helper 1 and 2 cytokines further supported a role for epigenetic mechanisms in T cell differentiation (Valapour *et al.* 2002).

1.3.2 Epigenetic regulation and the environment

Recent evidence supports the influence of both acute and chronic environmental stimuli on epigenetic modifications that can have lasting changes in gene expression and phenotype of organisms (Law & Jacobsen 2010). The relationship between the epigenome and environmental factors and its plasticity is an area of intense research both in humans and model organisms (Feil & Fraga 2012). Desirable expression landscapes via environmental epigenomic control can be seen in some natural phenomena. The early flowering reaction, or vernalisation, of plants in temperate climates following prolonged cold temperatures in winter is partly controlled by DNA methylation of the Flowering Locus C gene and a protein, VRN2 (Kim *et al.* 2009). The DNA methylome of the female honeybee (*Apis mellifora*) also undergoes significant changes upon receiving royal jelly as larvae to become fertile queens (Kucharski *et al.* 2008). Changes to the social environment of the bumble bee (*Bombus terrestris*) also resulted in DNA methylation and gene expression changes in COQ7, DNMT3 and foraging (*for*) and vitellogenin (*vg)* genes for both the queen and workers (Lockett *et al.* 2016).

The study of epigenetic control in mammals is still a developing area, however animal models show that nutrition and certain environmental exposure during development can lead to changes in the epigenome (Rosenfeld 2010). One of the most commonly referenced examples of this is in the agouti viable yellow (A^{vy}) allele in mice. When the retrotransposon intracisternal A-particle (IAP) is unmethylated *agouti* gene expression is changed and offspring produce a yellow coat in addition to obesity and diabetes (Wolff *et al.* 1998). Recently, hypermethylation of *GATA-4* in conjunction with vitamin A deficiency, has led to

the development of heart defects in rat offspring (Feng *et al.* 2013). A number of chemicals and pollutants such as tobacco smoke (Belinsky *et al.* 2002), particulate air pollution (Baccarelli *et al.* 2009), asbestos (Christensen *et al.* 2009), and bisphenol-A (BPA) (Dolinoy *et al.* 2007) have all been investigated in both humans and mice and are thoroughly reviewed by Feil & Fraga (2012).

Exercise and changes to the human epigenome has recently been a focus of research for both healthy and overweight subjects. Exercise intervention in twenty-three sedentary, but otherwise healthy males has been shown to change genome wide patterns of DNAm (Rönn *et al.* 2013). Using data generated with the HumanMethylation450 BeadChip they reported a wide range of altered CpG sites in type-2 diabetes candidate genes TCF7L2 and KCNQ1 with respective changes to mRNA expression. mRNA expression levels have previously been associated with exercise in an intensity dependent manner in skeletal tissue indicating that epigenetic changes observed at the whole genome level may be affecting metabolic processes (Egan *et al.* 2010; Rönn *et al.* 2013). Skeletal muscle displays a high degree of plasticity in its responses to environmental inputs and stressors that challenge the metabolic demands of the tissue. To investigate the effects of acute exercise on genes previously described to be differentially methylated in type-2 diabetes (Barrès *et al.* 2009). Barres and colleagues took skeletal muscle biopsies before and after intense exercise from a cohort of twenty-eight individuals. They reported that DNA methylation was unaltered following a 3-week training program, however RNA expression of target genes PGC-1α and TFAM was elevated, indicating a change in expression (Barres *et al.* 2012) . The authors concluded that acute exercise can lead to transient changes to DNA methylation and may include altered cytosine residues in other contexts such as; CpA, CpT or CpC that are not directly picked up when bisulphite sequencing (Barres *et al.* 2012).

1.3.3 Measuring epigenomic MABs

The most broadly useful methodology for epigenomic MAB quantification is high throughput DNA sequencing (HTS). HTS allows very precise measurement of diversity and identity of several nucleic acid types (Li & Stoneking 2012; Valdes *et al.* 2013). DNA methylation can be analysed by HTS after treatment of DNA with sodium bisulphite. This converts unmethylated cytosines to the RNA base uracil, but does not affect methylated cytosines. This process also converts 5-hydroxymethylcytosine (5hmC) and it is worth noting that all

methods that measure percentage CpG methylation based upon bisulphite conversion are actually measuring an aggregate of CpG methylation (5mC) and 5hmC. These two epigenetic marks are mediated by different enzymatic systems (Branco *et al.* 2012).

Future technologies and streamlined data analysis should allow differentiation between these two modifications so that any variable effects can be identified (Cavalcante *et al.* 2017; Clark *et al.* 2016). Analysis of HTS records methylated sites as 'C' and unmethylated cytosines as 'T'. Multiple molecules in the template pool are sequenced, so counts of 'C' and 'T' are a representation of the percentage methylation at each CpG site in that tissue, which corresponds to the proportions of cells where the site was methylated.

1.3.4 DNA methylation

DNA methylation (DNAm) is one of the best studied epigenetic marks and in its most common presentation involves the addition of a methyl group the 5' cytosine (5mC), of cytosine – guanine dinucleotides, hereafter referred to as CpGs (Jones *et al.* 2015). DNAm was first described as an epigenetic mark in 1975 by two key papers that proposed that DNA binding proteins could interpret DNAm marks and that methylation could silence gene expression and be inherited (Holliday & Pugh 1996; Riggs 1975). Vertebrate genomes can contain short (approximately 1kb) CpG rich regions known as CpG islands (CGI) and the majority of human research has focused on the role of DNA methylation near to transcription start sites (TSS, first exon) and promoter regions (Gardiner-Garden & Frommer 1987; Jones 2012). Recent technological advances have allowed the analysis of large numbers of CpG loci through bisulphite sequencing; this genomic data reinforces the importance of the position of 5mC in the gene of interest.

Analysis of DNA methylation by microarray and bead array screening technology has uncovered the role of several key genes involved in cellular senescence in mice (Maegawa *et al.* 2010) and humans (Bell *et al.* 2012; Branco *et al.* 2012). Age-related changes in DNA methylation have been used for developing a range of very precise MABs for humans (Bocklandt *et al.* 2011; Hannum *et al.* 2013; Horvath 2013; Weidner *et al.* 2014; Zbieć-Piekarska *et al.* 2015b). MABs for non-humans based on DNA methylation have been developed for humpback whales (Polanowski *et al.* 2014), and have also been tested on a small number of chimpanzees (Horvath 2013). The proportion of cells in a tissue that have a methylated or hydroxymethylated cytosine at a specific CpG site is referred to as percentage CpG methylation (Branco *et al.* 2012). Higher levels of CpG methylation generally cause

lower levels of transcription of mRNA from a given gene in vertebrates (Bell *et al.* 2012), but this pattern is gene specific and it can cause increased mRNA transcription from some genes (Smith & Meissner 2013).

The maintenance of methylation marks in mammals is carried out by DNA methyltransferases (DNMTs). DNMT1, DNMT3A and DNMT3B are all required for the initial establishment of methylation patterns in embryonic and early development and continuing established patterns of DNA methylation over the lifespan of the animal (Jones & Liang 2009). The mechanism for the removal of DNAm is less well understood, however active demethylation has been shown to occur through cell division or removal of the modified cytosine (Popp *et al.* 2010). Historic evidence suggested that DNAm of TSS in promoter CGI lead to silencing of expression, evidence now suggests that levels of gene expression are partly controlled by transcription factors (Jones 2012). Methylated promoter CGIs are generally restricted to those genes that are in a long-term repression state or in a state where expression would be undesirable. For example, genes that are expressed in germ cells are typically silenced in somatic cells. The theory that DNA methylation is not used as an initial silencing mechanism is supported by studies that show *de novo* methylation requires a nucleosome present (Ooi *et al.* 2007). Studies have tested the role of nucleosomes in recruiting DNMTs to lay down methylation and show that whilst active, TSSs are generally depleted of nucleosomes and lack a substrate for methylation (You *et al.* 2011).

Whilst the majority of DNAm research has focused on methylation of CpG loci due to its importance in human disease research, methylation of other cytosine dinucleotides such as CpA, CpT and CpC (non-CpG methylation) has been reported for many years but remains poorly understood (Patil *et al.* 2014). Non-CpG methylation was reported to be enriched in pluripotent cell types when compared with most differentiated cell types in humans, however it is not entirely restricted to these cells (Lister *et al.* 2009). Recent research has shown abundant methylation of this type in both human and mouse brain tissues (Varley *et al.* 2013; Xie *et al.* 2012). Measurements recording the effect of non-CpG methylation on *B29*, *PGC-1α* and *PDK4* gene expression provide the best evidence for a functional role for non-CpG methylation in mammals (Barrès *et al.* 2009; Malone *et al.* 2001). The establishment and maintenance of non-CpG methylation is thought to be linked to the continuous expression of DNMT3A, DNMT3B and DNMT3L, the latter of which does not have methyltransferase activity (Xie *et al.* 2012). Research of non-CpG methylation function in mammals may provide further opportunities for age biomarker identification in the future.

A large amount of research has focused on the role of DNAm and control of so called tumour suppressor genes (TSG), tumour suppressor microRNAs and DNA repair genes. It is generally supported that hypermethylation of these genes repress their transcriptional activity, leading to an inability to control cell growth, apoptosis, energy metabolism and proofreading errors (Lopez-Serra & Esteller 2012). Examples of expression regulation and disease pathogenesis by DNAm include protocadherin 17 (PCDH17) in gastric and colorectal cancers (Hu *et al.* 2013a) and PTEN in Cowden syndrome, where patients have and increased incidence of breast, thyroid and endometrial cancers (Hollander *et al.* 2011). Drugs targeting different parts of the epigenetic machinery can co-operate in the restoration of gene expression. 5-azacytidine and 5-aza-2'-deoxycytidine interferes with the activity of DNMT1 leading to genomic hypomethylation and reactivation of silenced TSG in humans (Esteller & Herman 2002). These agents have been widely used to demonstrate loss of methylation in specific gene regions and activation of those genes and has been thoroughly reviewed (Christman 2002).

Age-related DNAm analysis in humans and animals has been the focus of a number of studies following large array based research (Hannum *et al.* 2013; Horvath 2013). Methylation changes in some genes correlate very strongly with age in both array-based studies and specific age related gene research. Chapter 2 of this thesis entails a mini-review of the studies specifically investigating DNAm age correlations up to August 2017.

1.4 Avian genetics

Birds are the only extant descendants of dinosaurs and approximately 10,500 living species have been described, making them the most species-rich tetrapod group of vertebrates (Gill & Donsker 2013). Due to the vast amount of biological literature and data collected on birds they are often used as a model for investigating evolutionary and ecological hypotheses (Jetz *et al.* 2012). Commonly used as model species, the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*) have previously provided valuable insights into human disease and neuroscience (Glaze & Troyer 2006). Despite a clear case for a better understanding of bird evolution there remains a lack of annotated avian whole genomic data. A landmark study by Zhang *et al.* 2014 sought to address this gap in the data. Using a whole-genome shotgun approach the authors produced genomes for 45 species and in combination with the previously published species, these assemblies cover 92% of all avian orders (Wallis *et al.*

2004; Warren *et al.* 2010). Amongst amniotes, including fish and amphibians, birds have the smallest genomes. Mammals and reptiles have genomes that typically range in size from 1.0 to 8.2 Gb whilst avian genomes range from 0.91 to 1.3 Gb (Gregory 2001; Zhang *et al.* 2014). Compared to 24 mammal and three reptile genomes it was found that avian protein-coding genes were 50% and 27% shorter respectively. The authors report that this reduction was primarily due to intron deletions and reduced intergenic distances, resulting in a relatively higher gene density. Condensed genomes have also been observed in bats, the only flying mammalian group, and as such may represent an adaption for the rapid gene regulation that is required for powered flight in these animals (Organ *et al.* 2007; Zhang *et al.* 2014).

1.5 Avian tissues and DNA

A number of avian tissues can be effectively and ethically sampled from live animals including whole blood, feathers, faecal matter and buccal cells. Generally, blood would be taken from the birds' foot or wing, depending on size, species and age. Feathers can be taken from the breast or from other areas that do not negatively affect thermoregulation or flight. Faecal samples can also be collected upon direct observation, but can be challenging for burrowing animals (McInnes *et al.* 2017). Single use buccal swabs can also be used for captured animals but may require multiple swabs to yield sufficient DNA for downstream use.

When whole blood extraction is difficult or deemed too invasive, the use of feathers can simplify the sampling of genomic DNA from birds. Feathers consist of a calamus (quill), which extends to the rachis (main shaft), which then supports the feather barbs (Prum 1999). Using feathers generally minimises the stress placed on the bird and simplifies the sampling procedure, particularly for larger birds. Bello *et al.* (2001) describes a technique for the isolation of high quality genomic DNA from feather quill tips. Briefly, a 0.5 – 1 cm section of quill was cut from the feather and lysed, larger feathers were incubated overnight and DNA was purified with phenol:cholorform:isoamyl alcohol. This protocol was done on samples from over 800 birds covering 120 species, and had yields ranging from 1 – 100 μg (Bello *et al.* 2001). They also report successful PCR amplification for chromodomain-helicase-DNA-binding protein (CHD) gene based sex identification for non-ratite birds (Ellegren 1996; Jensen *et al.* 2003). Feather barbs have also been used to obtain mitochondrial DNA (mtDNA) for species identification methods. Speller *et al.* (2011) was able to improve upon previous studies by extracting mtDNA from barbs sourced from 'fresh' feathers, historical

samples and archaeological samples (Rawlence *et al.* 2009). The authors report PCR amplification for Cytochrome b yielded similar successful results for the North American wild turkey (*Meleagris gallopavo*) and Canada goose (*Branta canadensis*), and that follow up sequencing yielded results that were similar to GenBank references. The authors concluded that mtDNA is recoverable from feather barbs rather than the entire feather, and can be used for downstream sequencing applications (Speller *et al.* 2011). Previous research has also shown that feathers used for PCR based molecular sex determination yielded complete agreement with matched blood samples from 102 individual Black-capped chickadees (*Poecile atricapilla*) indicating that non-destructive methods of sample collection can be widely used for molecular methods (Harvey *et al.* 2006).

There is ongoing discussion regarding the impact of invasive sampling techniques including blood and feather sampling. Previous studies have shown that blood sampling of Cliff swallows (*Petrochelidon pyrrhonota*) led to a 33% reduction in annual survival (Brown & Brown 2009). For very young nestlings, taking blood could adversely affect survival depending on the volume taken, and for unfeathered young there are fewer alternatives to blood. Additionally, faecal samples can be too degraded to allow for successful genomic applications or contain contaminants that hinder molecular techniques. One alternative to the described sampling techniques are buccal swabs, these are routinely used in forensics and pathology (Handel *et al.* 2006; Walsh *et al.* 1991). A study comparing buccal samples and blood samples from the Common swift (*Apus apus*) reported that a PCR product for sex determination was achieved in 89% of buccal samples (N = 53 nestlings) and 98% of these were identical to sexes determined by the matched blood sample (Wellbrock *et al.* 2012). This result was further supported by a previous study that reported an 82.2% (N = 107) PCR success rate for avian buccal sexing in the Great cormorant (*Phalacrocorax carbo*) (Arima & Ohnishi 2006).

1.6 Ageing and seabirds

Seabird population dynamics can be used to quantify simple binary and long-term changes in ecosystems and certain population responses require some information on the standing age-structure of the population under study (Dale & Beyeler 2001; Parmesan 2006; Piatt *et al.* 2007). The chronological age of an animal is a critical factor for many biological processes that can change with time. Age can be a useful marker of life history events including age of reproductive maturity, reproductive frequency and breeding success (Bradley & Safran 2014;

Charpentier *et al.* 2008; Jarman *et al.* 2015). In animal populations, age-class distribution is both a determinant of the current growth rate and a reflection of past growth rates and may include the previous effects of harvesting or other human influences (Ozgul *et al.* 2010; Polanowski *et al.* 2014; Tkadlec & Zejda 1998).

There are currently many useful methods for age estimation in wild animals, however some are only applicable to dead animals, highly invasive, or are not currently sensitive enough for the needs of specific projects (Campana 2001). The majority of seabirds have no easily identifiable marks to age individuals, limiting the amount of demographic data that can be obtained.

One method of ageing birds uses pentosidine, a product of nonenzymatic glycation, also known as an advanced glycation end product (AGE), that accumulates in animal tissues. Previous studies have reported a linear association of pentosidine concentration and chronological age in several bird species (Cooey *et al.* 2010; Dorr *et al.* 2017; Fallon *et al.* 2006). The first of these studies in wild birds identified a linear relationship ($p < 0.001$) with age of pentosidine from foot webbing samples from dead California gulls (*Larus californicus*) (Chaney Jr *et al.* 2003). This linear relationship was confirmed in Double-crested cormorants and was able to correctly identify breeding status (83.5%) in modelled classifications (Dorr *et al.* 2017). Generally, these observations support the error, damage and cross-linking theories of ageing, whereby cells can be damaged by accumulation of cross-linked proteins (Bjorksten 1968; Hayflick 1985). However, no association between chronological age and pentosidine concentration was reported in Common gulls (*Larus canus*). This result is surprising given previous research in other bird species but may indicate an ability to abrogate pentosidine accumulation (Rattiste *et al.* 2015). Results from a recent thesis also did not identify a correlation with skin pentosidine and age in Bridled terns (*Onychoprion anaethetus*). The author indicated that low collagen levels in the skin may result from diet or a life at sea, but did not rule out a possible molecular mechanism of resistance to oxidative damage (Labbé 2017). Consequently, a minimally invasive and non-lethal MAB is required for collecting vital life-history information in long-lived seabirds.

1.7 Short-tailed shearwater

The Short-tailed shearwater (STS, described as *Ardenna tenuirostris* (Christidis & Boles 2008; Penhallurick & Wink 2004); originally *Puffinus tenuirostris* (Temminck 1835)), also

known as the ‘mutton bird’, is the most abundant seabird species in Australian waters and are one of the few native Australian birds still actively harvested. There are approximately 280 colonies located around the south east coast of Australia with the largest in the world (estimated at 2.8 million pairs), found on Babel Island. Of these colonies, 160 are found on the island of Tasmania with an estimated 11 million burrows.

Shearwaters breed in Australia and each year migrate distances upwards of 15,000 km northwards to coastal areas around the Aleutian Islands, Kamchatka Peninsula and Japan (Marshall & Serventy 1956). Shearwaters represent an ideal model organism for the development of a seabird-ageing assay, as I describe here.

As the shearwater is currently listed by the International Union for the Conservation of Nature as ‘least concern’, is abundant and routinely monitored in both Tasmania and Japan, additional sampling for this study can be carried out during normal handling (IUCN 2012). Known age cohorts, such as the Fisher Island population (Figure 1), are critical for the development of future ageing assays to ensure accurate estimations are made (Bradley *et al.* 1991). Burrows on Fisher Island are monitored twice yearly and known-age individuals are often resighted at the same or nearby burrows (Figure 2). Model validation and age estimates can then be carried out in another colony of unknown age birds. Molecular evidence from mitochondrial cytochrome *b* sequencing suggests that the STS is closely related to the Sooty shearwater (*Ardenna griseous*) and the Great shearwater (*Ardenna gravis*), with all three species in order Procellariiformes (Penhallurick & Wink 2004). Investigating sequence conservation between species and higher levels of organisation is important as it may lend support to the use of the same age model across species with minimal calibration required.

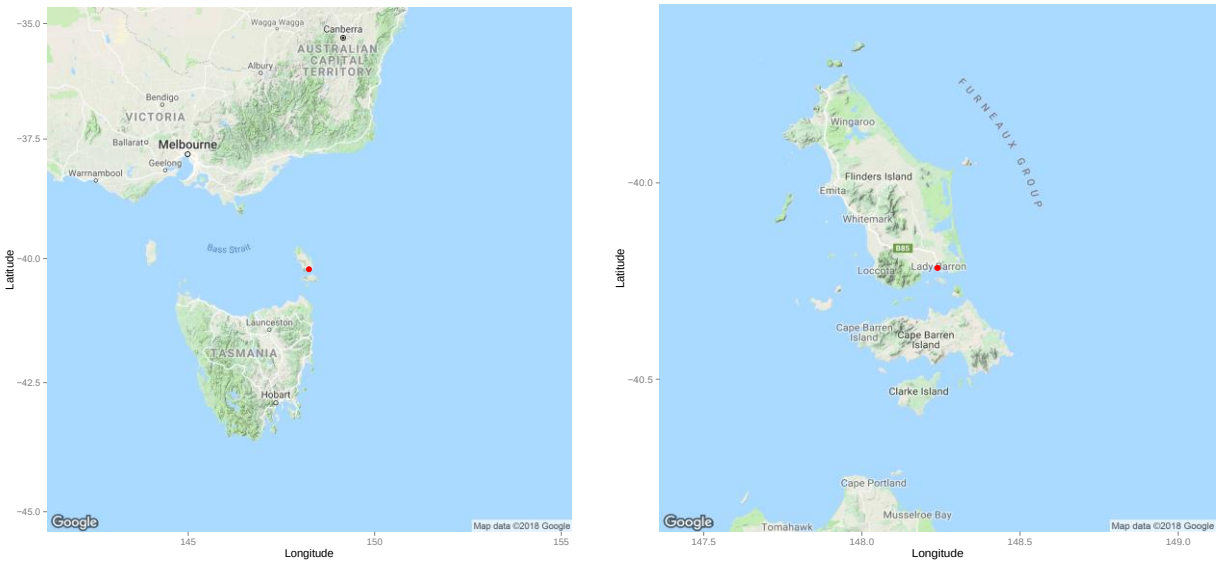


Figure 1 | Fisher Island field site location maps. These maps show Fisher Island in the context of **A.** Bass Strait and **B.** Furneaux Island Group.

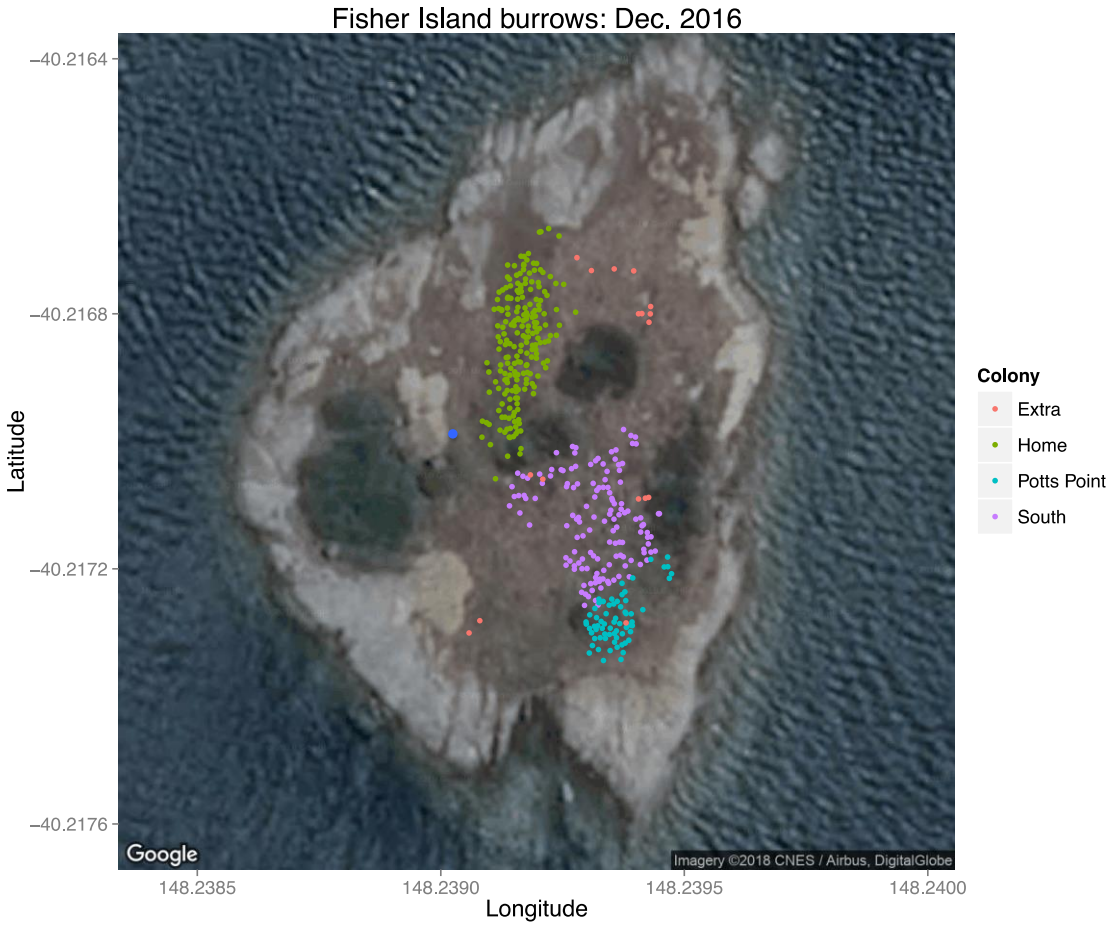


Figure 2 | Fisher Island approximate burrow locations (2016). The entirety of Fisher Island is shown here with the field hut (blue dot) and burrows shown separated into locally distinct 'colonies' labeled as 'Home', 'South' and 'Potts Point' in green, purple and cyan respectively. External or uncertain/temporary burrows around these areas are shown in red.

1.8 Thesis aims and overview

To our knowledge, the avian methylome remains largely unexplored. The aim of this thesis research was to determine if DNA methylation measured in bird tissues could provide a reliable estimate of individual age. We hypothesised that recent breakthroughs in mammal age estimation would also be applicable as molecular age biomarkers in a long-lived seabird. To achieve these aims, I used the Short-tailed shearwater as a model species to:

1. Investigate age-related genes identified in mammalian species in shearwater blood and feather DNA samples.
2. Quantify DNAm and identify age-related CpG sites in a restricted whole-epigenome analysis of known age shearwater blood samples.
3. Further characterise age-related genes and estimate chronological age in previously unbanded shearwater individuals.

In chapter 3 of this thesis we established tissue specific processing and DNA extraction techniques in the laboratory and determined optimal field protocols for collection of seabird samples for age related work. We also quantified shearwater DNAm in a target set of age-related genes previously identified in mammals. In chapter 4, we carried out a restricted epigenome-wide MAB search in shearwater blood samples. In chapter 5, we identified the genes associated with age in the shearwater and used these to estimate age in populations of birds not previously monitored or banded. Chapters 2 – 4 were prepared as separate manuscripts for publication and therefore may contain minimal repetition between chapters. At submission, chapters 2, 3 and 4 have been published, and we plan to submit chapter 5 for review.

Chapter 2 - Measuring animal age with DNA methylation: from humans to wild animals

Published as: De Paoli-Iseppi R, Deagle BE, McMahon C, Hindell MA, Dickinson JL and Jarman SN (2017) Measuring animal age with DNA methylation: from humans to wild animals. *Frontiers in Genetics: Genetics of Ageing*, 8, 106.

2.1 Abstract

DNA methylation (DNAm) is a key mechanism for regulating gene expression in animals and levels are known to change with age. Recent studies have used DNAm changes as a biomarker to estimate chronological age in humans and these techniques are now also being applied to domestic and wild animals. Animal age is widely used to track ongoing changes in ecosystems, however chronological age information is often unavailable for wild animals. An ability to estimate age would lead to improved monitoring of (i) population trends and status and (ii) demographic properties such as age structure and reproductive performance. Recent studies have revealed new examples of DNAm age association in several new species increasing the potential for developing DNAm age biomarkers for a broad range of wild animals. Emerging technologies for measuring DNAm will also enhance our ability to study age-related DNAm changes and to develop new molecular age biomarkers.

2.2 Introduction

Biological ageing involves complex interactions of accumulating organ, cellular and DNA damage leading to functional decline and increased risk of death (Fontana *et al.* 2010). Biological functions including the age of reproductive maturity (Charpentier *et al.* 2008; Jones *et al.* 2014), reproductive frequency (Froy *et al.* 2013), and mortality (Pérez-Barbería *et al.* 2014) can be better understood in the context of age. Estimates of chronological age are therefore useful for understanding these key ecological characteristics of wild animals. However, chronological age is difficult to estimate in individuals of most animal species (Nussey *et al.* 2013). Many species lack measurable external features that change with age. Therefore, new methods that allow estimation of chronological age will enhance our understanding of ageing and population biology in wild animals.

DNA methylation (DNAm) at cytosine guanine dinucleotides (CpGs) is the best studied epigenetic modification and can repress gene expression when associated with gene promoters (Jones *et al.* 2015). CpG methylation in mammals is regulated by DNA methyltransferases (DNMTs). DNMTs are required for the initial establishment of methylation patterns in early development (Law & Jacobsen 2010); and for maintaining established patterns of DNAm over the lifespan of the animal (Jones & Liang 2009). There are two types of age-associated DNAm in vertebrates, ‘epigenetic drift’ and ‘clock-type’ DNAm (Jones *et al.* 2015). DNMT1 is primarily responsible for maintaining CpG

methylation and its decline in activity with age is thought to contribute to a decrease in global methylation or 'drift' in ageing cells (Jones *et al.* 2015). However, gene-specific DNAm change with age may be regulated by other de-novo DNMTs, such as DNMT3b (Lopatina *et al.* 2002). 'Clock-type' age-associated DNAm is a change in methylation proportion (either an increase or decrease) at specific CpG sites. Changes at clock-type CpGs may be related to functional changes in gene expression with age (Horvath 2013; Steegenga *et al.* 2014).

In this mini-review, we summarise current knowledge of observed age-related changes of CpG DNAm in mammals, reptiles, birds and fish. Recent technological advances have enabled the relatively quick analysis of large numbers of CpG loci (Parle-Mcdermott & Harrison 2011). This has greatly increased the number of studies that have observed age-related DNAm in humans and model organisms (Jarman *et al.* 2015). We also describe how this new information can be used to develop molecular age biomarkers (MABs) for non-model animals. We explore environmental and behavioural studies of DNAm that are relevant to age estimation. We also discuss emerging technology and their potential for application in wild animals.

2.3 DNA methylation ageing signals in humans and mice

Age estimation models based on CpG DNAm combine information from CpG sites that have the highest correlation of DNAm levels with age. These are calibrated using tissue samples from known-age individuals. Early ageing models for humans used single tissues and small numbers of CpG sites (Bocklandt *et al.* 2011). Recent more precise models predict age from multiple tissues (Horvath 2013). DNAm changes with age can contribute to altered gene expression levels during normal ageing (Zykovich *et al.* 2014) and disease (Nilsson *et al.* 2014). Therefore, the effect of CpG DNAm on transcriptional regulation of gene expression has been the focus of intense research (Goyns 2002) (Figure 1). The first studies to identify human age-related DNAm changes studied monozygotic twins, where epigenetic drift with age was observed when comparing older and younger twins (Fraga *et al.* 2005). This observation was supported by further studies of monozygotic twins and healthy controls that led to the first epigenetic age models (Boks *et al.* 2009). A detailed summary of age-related DNAm studies is shown in Table S1.

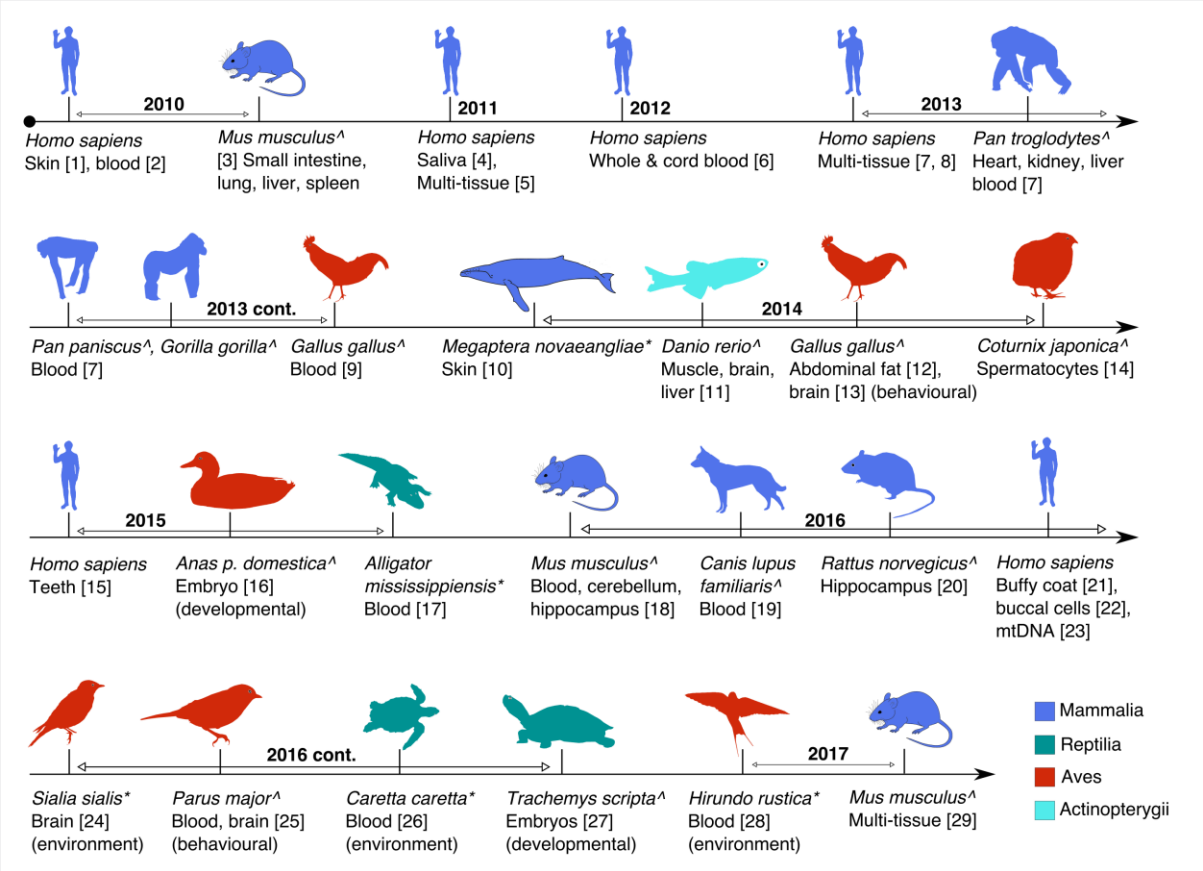


Figure 1 | Timeline of the major studies and tissues analysed for global or targeted DNA methylation in this review. Studies are age-associated except where indicated. Superscripts: ^(captive raised or model studies) and *(wild animal studies). (1, Grönniger et al., 2010; 2, Teschendorff et al., 2010; 3, Maegawa et al., 2010; 4, Bocklandt et al., 2011; 5, Koch and Wagner, 2011; 6, Garagnani et al., 2012; 7, Horvath, 2013; 8, Hannum et al., 2013; 9, Gryzinska et al., 2013; 10, Polanowski et al., 2014; 11, Shimoda et al., 2014; 12, Sun et al., 2014; 13, Nätt et al., 2014; 14, Andraszek et al., 2014; 15, Bekaert et al., 2015; 16, Yan et al., 2015; 17, Nilsen et al., 2016; 18, Spiers et al., 2016; 19, Gryzinska et al., 2016; 20, Penner et al., 2016; 21, Christiansen et al., 2016; 22, Eipel et al., 2016; 23, Mawlood et al., 2016; 24, Bentz et al., 2016; 25, Verhulst et al., 2016; 26, Caracappa et al., 2016; 27, Matsumoto et al., 2016; 28, Romano et al., 2017; 29, Stubbs et al., 2017).

Age-related DNAm changes have been identified in several human and mouse tissues. Analysis of human skin samples for DNAm changes identified a set of CpG sites that were affected by chronological age. Thirty sun-exposed and sun-protected skin biopsy samples were analysed using the Infinium HumanMethylation27 (27K) BeadChip (Illumina). The study identified 104 CpG sites that had a DNAm relationship with age, but not by sun exposure (Grönniger *et al.* 2010). Another example of age-related CpG DNAm was found in the saliva samples of 34 identical twins (21 – 55 years old). This study identified 88 CpG sites where DNAm levels were significantly correlated with chronological age (27K BeadChip) (Bocklandt *et al.* 2011). The genes identified were involved in age-related cardiovascular and neurological diseases (Bocklandt *et al.* 2011; Park *et al.* 2007). In mice, linear age-related methylation was identified in multiple CpG sites from 12 different genes measured in intestine, lung, liver and spleen. This indicated that similar age-related changes in DNAm levels can be found in humans and other mammals (Maegawa *et al.* 2010) .

Most clock-type DNAm changes are tissue specific, however several studies have investigated age-associated changes in multiple tissues. Non cell-type dependent DNAm changes were identified in a study that combined several published CpG DNAm data sets to predict age using multiple tissues (Koch & Wagner 2011). Four CpG sites in *TRIM58*, *KCNQ1DN*, *NPTX2* and *GRIA2* were identified from a set of 431 hypermethylated CpGs. Multiple linear regression of the DNAm levels for each CpG site against the known donor age resulted in a model with a mean absolute difference (MAD) of ± 10.3 years (Koch & Wagner 2011). Another multi-tissue model based on 353 CpGs was developed from 82 publicly available data sets of 8,000 healthy tissue and cell types and had a median absolute difference of ± 3.6 years (Horvath 2013). This study predicted age using a greater range of tissue types and show that cancer can lead to an increased age as measured by DNAm. The model was more accurate in heterogeneous tissues such as blood and saliva compared to breast tissue, dermal fibroblasts and skeletal muscle. The author suggested hormonal or cancer effects as possible causes of this variation. Horvath (2013) proposed that this model measures the cumulative work of an epigenetic maintenance system that likely functions over the entire mammalian lifespan. The DNA methylome and human ageing rates were also compared using a far greater number of CpG sites, the HumanMethylation450 (450K) BeadChip (Hannum *et al.* 2013). Here, a MAB using 71 CpGs predicted chronological age with a MAD of ± 3.9 years. In mice, a multi-tissue age predictor has been developed using 329 CpG sites, giving a median absolute error of ± 3.3 weeks (Stubbs *et al.* 2017). Together,

these three models show that methods based on large numbers of markers can improve precision of age estimates in model organisms or humans.

Lifestyle factors can influence DNAm levels in humans and mice and could impact chronological age estimation if not corrected for. A positive difference between estimated DNAm age and known age suggests that an individual is biologically older than their chronological age. Described epigenetic clocks have been used to show decreased age acceleration dependent on diet (Quach *et al.* 2017) and increased age acceleration associated with smoking (Gao *et al.* 2016; Zaghlool *et al.* 2015) and body mass index (Horvath *et al.* 2014). Recent research has also highlighted the differences between human and murine DNAm clocks (Wagner 2017) and the effects of calorie restriction on mouse biological age (Petkovich *et al.* 2017).

The use of DNAm biomarkers for specialised forensic applications generally involves using fewer CpG sites in simpler assays. A model based on three CpG sites in human blood yielded a MAD from known-age samples of ± 5.4 years, which was an improvement over other non-epigenetic molecular ageing techniques (Weidner *et al.* 2014). In two studies, a small number of CpG sites in one gene region (*ELOVL2*) allowed simplification of technical analyses while maintaining prediction accuracy (MAD ± 3.9 years) (Zbieć-Piekarska *et al.* 2015a; Zbieć-Piekarska *et al.* 2015b). Single multiplex reactions such as methylation-sensitive single-nucleotide primer extension can be used to make age biomarkers that are cheaper to run than pyrosequencing or microarray assays. For example, one study using this method with eight CpG sites led to age predictions with a MAD of ± 6.07 years (Vidal-Bralo *et al.* 2016). While this approach had lower precision than previous studies it is still a feasible tool for estimating age using adult blood. One study implemented the models published by both Horvath (2013) and Hannum et al. (2013). Here, buffy coat was isolated from twins (30 – 82 years) and age was predicted (Christiansen *et al.* 2016). This resulted in MADs of ± 5.6 years for the 353 CpG Horvath model and ± 5.4 years for the 71 CpG Hannum model demonstrating the benefit of using a higher number of CpG sites.

Methylation changes in mitochondrial DNA (mtDNA) associated with age have been identified in humans. An epigenetic model based on two CpGs with a MAD of ± 9.3 years was developed from the blood of 82 individuals (18 – 91 years). Age was correlated with mtDNAm at two sites, M1215 and M1313, in the 12s *MT-RNR1* gene (Mawlood *et al.* 2016). mtDNA overall has a low level of CpG methylation (2-6%), so detection of age-related CpG

levels required high assay precision. This is the only study of age-related mtDNA methylation and it is still uncertain whether mtDNAm-based models will match the accuracy of models using genomic biomarkers.

2.4 Quantifying DNA methylation, environmental effects and age in model and wild animals

2.4.1 Mammals

DNAm age biomarkers have only been developed for a small number of wild mammal species. In long-lived species, obtaining known-age calibration sample sets that cover the entire lifespan is a significant obstacle. DNAm age estimation is most advanced in species closely related to humans. The age of chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*) and gorillas (*Gorilla gorilla*) were estimated using the 353 CpG clock MAB created for humans (Hernando-Herraez *et al.* 2013; Horvath 2013; Pai *et al.* 2011). Results from blood samples showed that in both chimpanzees and bonobos the model had an accuracy similar to that found in humans; however, accuracy was reduced in gorillas (Horvath 2013).

Humpback whales (*Megaptera novaeangliae*) have successfully been used as a test case for applying knowledge of human age-related clock type DNAm change to estimate age in a long-lived wild mammal. 45 known-age samples were used to calibrate a DNAm age model. Seven of 37 CpG loci screened by pyrosequencing showed significant age-related DNAm. The three sites with the strongest relationship with age were used to predict whale age from skin with a MAD of 3.75 years. This model also predicted the correct order of ages in samples with known kinship in more than 93% of cases (Polanowski *et al.* 2014).

Global DNAm levels in dogs change with age as a result of epigenetic drift. Significant differences in relative global DNAm levels have been found amongst pups (43.5%), adolescents (53.6%), adults (61.5%) and old dogs (81.2%) (Gryzinska *et al.* 2016). Clock-type DNAm age biomarkers based on multiple CpG sites have been developed for dogs. These models were calibrated using blood from multiple known-age animals and could predict age with a minimum MAD of 23.1 months (Ito *et al.* 2017).

2.4.2 Birds

DNAm patterns in birds are relatively unexplored compared to mammals (Head 2014). Most avian DNAm research focuses on chickens (*Gallus gallus*) and quails (*Coturnix japonica*). Observation in *G. gallus* of unmethylated CpG islands in gene promoters (Li *et al.* 2011) and altered CD4 gene transcription due to increased DNAm of the promoter, indicates a similar regulatory function to that in mammals (Luo *et al.* 2011) (Figure 2). An age-related decrease in percentage DNAm of six CpG sites in the *PPARγ* promoter in 2, 3 and 7-week-old *G. gallus* has also been reported (Sun *et al.* 2014). Global DNAm levels have been shown by immunoenzymatic assay to decrease with age in *G. gallus* (Gryzinska *et al.* 2013). Higher global DNAm levels of 55-week-old hens (*G. gallus*) compared to 20-week-old individuals have recently been shown in breast tissue using whole-genome bisulphite sequencing. Of 2,714 identified differentially methylated regions, 378 were mapped to gene promoters including *ABCA1*, *COL6A1* and *GSTT1L*. CpG sites in these genes were hypermethylated with age and could be used for future age biomarker studies (Zhang *et al.* 2017). DNAm analysis of *C. japonica* DNA by gel imagery showed that 15-week-old quails had increased DNAm of the *RN28S* gene compared to 52-week-old individuals (Andraszek *et al.* 2014).

DNAm studies of behavioural traits in several bird species could yield potential targets for age biomarker development if the identified genes are linked to behavioural change during an animal's life. The DNAm level in the dopamine receptor D4 (*DRD4*) gene in great tits (*Parus major*) was shown to be associated with variations in exploratory behaviour (Verhulst *et al.* 2016). Methylation tiling arrays have linked the promoters of the zinc finger RNA binding protein (*ZFR*) gene and male hypermethylated region (*MHM*) with sex dependent gene expression in chicken brain samples (Nätt *et al.* 2014). Several of the differentially expressed genes were known to affect behaviours including exploration and fearfulness.

Age-associated global DNAm could be affected by environmentally altered DNMT expression in birds. Increases in mRNA for DNMT1, DNMT3A and methyl binding protein MBD5 were found after a 1 °C increase in incubation temperature in several tissues of embryonic Peking ducks (*Anas platyrhynchos domestica*). These changes in expression levels of methylation-interacting enzymes could lead to overall changes in DNAm of bird genomes with age (Yan *et al.* 2015). A single CpG in the *ERα* promoter in wild eastern bluebirds (*Sialia sialis*) was positively correlated with yolk testosterone concentration and nestling growth rate.

Figure 2. Variable global methylation in vertebrates	Genome size (Mb)*	Assembly status*	GC%	5mC	5mC/GC proportion	Methylation pattern	CpG island hypomethylation	Gene-body methylation	DNMTs (3A/B)	DNMT3L (imprinting)	Ref
Fish				1.70^	3.59	Global	-	-	Yes	-	8, 9
Cyprinus carpio (common carp)	1546.8	Full	37.1	1.38	3.72						5
Danio rerio (zebrafish)	1427.2	Full	36.7	1.27	3.46		Yes	Yes	Yes	No	5, 3, 7
Mammals				0.88^	2.07	Global	Yes	Yes	Yes	Yes	8, 9
Canis lupus familiaris (dog)	2254.6	Full	41.1	0.67	1.63						4
Balenoptera physalus (fin whale)	-	-	41.3	0.94	2.28						4
Mus musculus (mouse)	2671.8	Full	42.4	0.95	2.24						2, 3
Homo sapiens (human)	2996.4	Full	40.9	0.88	2.15						1
Macaca mulatta (rhesus monkey)	3033.6	Full	41.5	0.86	2.07						2
Reptiles				Variable[6]	-	Global	Yes	Yes	Yes	-	8, 9
Snakes & lizards				1.11	3.13						
Iguana iguana (green iguana)	-	-	44.3	1.36	3.07						6
Thamnosophis lateralis (lateral water snake)	-	-	44.3	1.41	3.18						6
Turtles & alligators				0.96	2.20						
Chelonia mydas (green sea turtle)	2208.4	Scaffold	43.7	1.00	2.29						6
Chrysemys picta (painted turtle)	2365.7	Chromosome	44.5	0.96	2.16						6
Alligator mississippiensis (american alligator)	2161.7	Full	44.3	0.96	2.17						6
Birds				1.02^	2.41	Global	Yes	Yes	Yes	-	8, 9
Rhea americana (greater rhea)	-	-	44.0	0.98	2.23						1
Columba livia (rock pigeon)	1063.0	Full	41.4	1.02	2.46						1
Corvus corone (crow)	-	-	45.4	1.15	2.53						1
Gallus gallus (chicken)	1230.2	Chromosome	42.9	1.04	2.42				Yes	No	5, 3

Figure 2 | Variable global methylation in vertebrates. *Current genome on NCBI (if available). ^Average 5 mC for classes (Jabbari et al., 1997). (1, Ehrlich et al., 1982; 2, Gama-Sosa et al., 1983; 3, Yokomine et al., 2006; 4, Jabbari et al., 1997; 5, Vanyushin et al., 1970, 1973; 6, Varriale and Bernardi, 2006; 7, Shimoda et al., 2014; 8, Ponger and Li, 2005; 9, Okamura et al., 2010; Tree, Letunic and Bork, 2006).

While DNAm at this CpG appears to depend on maternal and environmental conditions, it may help to better estimate age in pre-fledgling chicks (Bentz *et al.* 2016).

The epigenetic effect of adverse environmental conditions has recently been analysed in a wild barn swallow (*Hirundo rustica*) population. After particulate matter exposure, DNAm levels in two *Clock* gene loci were significantly increased in chicks (7 – 5 days old) and mothers. This study is a good example of a targeted DNAm approach and is the first study to show that DNAm levels can change in response to anthropogenic pollutants in wild birds (Romano *et al.* 2017). Exposure to DNAm altering compounds could be used to produce DNAm age biomarkers if exposure is consistent over time and among individuals in a population.

2.4.3 Reptiles

Reptile DNAm has not been well studied in general and there is little data on age-related DNAm. Reptilian CpG island positions relative to promoters are similar to those found in mammals and birds, indicating that DNAm has a similar regulatory function (Head 2014; Varriale & Bernardi 2006). Adult American alligators (*Alligator mississippiensis*) have consistently lower global DNAm than sub-adults and captive juveniles (Parrott *et al.* 2014). The decrease in global DNAm through epigenetic drift is consistent with that found in all other studies to date (Nilsen *et al.* 2016). While there are no age-related clock-type reptilian studies, some have measured changes in DNAm due to environmental influences. Reduced global DNAm due to phenotypic differences was found in the loggerhead sea turtle (*Caretta caretta*) (Caracappa *et al.* 2016). In the red-eared slider turtle (*Trachemys scripta*) CpG DNAm levels of the *aromatase* gene were associated with shifts in egg incubation temperature (Matsumoto *et al.* 2013; Matsumoto *et al.* 2016).

2.4.4 Fish

Fish DNAm is better studied compared to birds and reptiles, with the majority of research on the zebrafish model (*Danio rerio*). Zebrafish embryos show high levels of global CpG methylation (80%), which is similar to mouse (74%) and also have depletion of methylation around transcriptional start sites similar to mammals (Feng *et al.* 2010). However, there are important differences in epigenetic reprogramming during early embryogenesis that are reviewed elsewhere (Head 2014; Potok *et al.* 2013) (Figure 2). A gradual and clear loss of CpG DNAm using methylation-sensitive enzymes and cloning was shown in 3, 18 and 30-month-old zebrafish (Shimoda *et al.* 2014). As with other non-model animal groups, there is little DNAm age data for wild fish.

2.5 Future directions

Changes to DNAm patterns have been established as biomarkers of chronological and biological ageing in humans and have great potential for age estimation in wild animals. Global DNAm hypomethylation correlating with age has been found in a range of wild animals, suggesting that epigenetic drift occurs in most vertebrates. Numerous clock-type age-related CpG sites have also been identified, several of which appear to be well conserved

in mammals (Table S1). Both epigenetic drift and clock-type DNAm changes could be used for age estimation in vertebrates.

New technologies for measuring changes in DNAm relating to age will improve our ability to generate age biomarkers. Nanopore technology has recently improved so that it is possible to identify cytosine and adenosine methylation variants in *E. coli* (Rand *et al.* 2017). An advantage of nanopore technology is that relatively small amounts of non-treated DNA are required to produce long sequence reads compared to bisulphite treated DNA. This may be advantageous for animal studies where DNA yield from the target tissue is low, such as feather quill ends (Simpson *et al.* 2017). Digital restriction enzyme analysis of methylation (DREAM) allows the precise measurement of CpG sites in a global context (Jelinek *et al.* 2012; Maegawa *et al.* 2014). Preparation of samples for both methods is relatively simple and does not require a reference sequence to identify age related signals (Jelinek & Madzo 2016).

Measuring wild animal age with DNAm has diverse applications in ecological and environmental research. Populations of known-age wild animals will be particularly important for this type of research. Age estimates generated from a robust DNAm model could be used to understand survival, reproductive potential and biological ageing (Jarman *et al.* 2015; Jazwinski & Kim 2017). Development of age biomarkers in new species will benefit immensely from information on age-related DNAm change gathered from humans and model organisms. This field is poised to change the way that the age of wild animals is determined.

Chapter 3 - DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird

Published as: De Paoli-Iseppi R, Polanowski AM, McMahon C, Deagle BE, Dickinson JL, Hindell MA and Jarman SN (2017) DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. PLoS One, 12, (12), e0189181.

3.1 Abstract

Most seabirds do not have any outward identifiers of their chronological age, so estimation of seabird population age structure generally requires expensive, long-term banding studies. We investigated the potential to use a molecular age biomarker to estimate age in short-tailed shearwaters (*Ardenna tenuirostris*). In this chapter I quantified DNA methylation in several *A. tenuirostris* genes that have shown age-related methylation changes in mammals. In birds ranging from chicks to 21 years of age, bisulphite treated blood and feather DNA was sequenced and methylation levels analysed in 67 CpG sites in 13 target gene regions. From blood samples, five of the top relationships with age were identified in *KCNC3* loci (CpG66: $R^2 = 0.325$, $p = 0.019$). In feather samples *ELOVL2* (CpG42: $R^2 = 0.285$, $p = 0.00048$) and *EDARADD* (CpG46: $R^2 = 0.168$, $p = 0.0067$) were also weakly correlated with age. However, the majority of markers had no clear association with age (of 131 comparisons only 12 had a p-value < 0.05) and statistical analysis using a penalised lasso approach did not produce an accurate ageing model. Our data indicate that some age-related signatures identified in orthologous mammalian genes are not conserved in the long-lived short tailed shearwater. Alternative molecular approaches will be required to identify a reliable biomarker of chronological age in these seabirds.

3.2 Introduction

Ageing is generally considered to involve the complex interaction of accumulating organ, cellular and DNA damage which lead to functional decline and increased risk of disease and death (Fontana *et al.* 2010). Whilst ageing and senescence have now been observed in many wild animals, the rate of reproductive and functional decline varies (Hayflick 2007; Nussey *et al.* 2013). The effect of increasing chronological age is associated with key ecological characteristics including reproductive maturity (Charpentier *et al.* 2008; Essington *et al.* 2001; Nussey *et al.* 2008), reproductive frequency (Bradley & Safran 2014), and mortality (Pérez-Barbería *et al.* 2014). Given that many species lack measurable external changes with age, estimating chronological age is often difficult. Developing a non-lethal method for the accurate estimation of chronological age is an important first step for furthering our understanding of ageing in wild animals.

Within animal populations, age structure is a significant determinant of population growth rate and can reflect past population growth rates, for example, an over-representation of young individuals in fast-growing populations (Ozgul *et al.* 2010; Tkadlec & Zejda 1998). Age has been associated with maturation, litter size, offspring body size and survival (Acker *et al.* 2014; Massot *et al.* 2011). Age structure can also be important for the management of wild populations including population viability studies for endangered species (Beissinger & Westphal 1998). With the recent rapid advance of miniature sensors, free-living animal research is increasingly turning to the remote recording of activity and associated environmental conditions, which in combination with age information, could lead to a much broader view of animal communities (Wilmers *et al.* 2015). Questions about avian physiological performance, foraging location and migration also require accurate age data to understand age-specific behaviours (Shepard *et al.* 2011; Watanabe *et al.* 2014; Wilmers *et al.* 2015).

Molecular biomarkers have shown promise for age estimation in long-lived animals (Jarman *et al.* 2015). Chronological age is not always a reliable indicator of overall bodily function, or 'biological age', because individuals experience different exposure to damaging elements throughout life and their ability to repair damage varies (Levine 2013; Selman *et al.* 2012). Several of the molecular changes involved in ageing can be quantified and these are an area of intense study in humans (Levine 2013). Recent evidence suggests that specific human behaviours associated with disease such as smoking (Zaghlool *et al.* 2015) and alcohol consumption (Liu *et al.* 2016; Weng *et al.* 2015) can be identified by quantifying epigenetic modifications. Whilst biological age has potential as a biomarker, we did not use damage-induced markers due to their innate difference from chronological age. Telomere restriction fragment (TRF) lengths have previously been shown to correlate with age in some mammals (Ou *et al.* 2012; Polanowski *et al.* 2014). TRF analysis has provided a good opportunity to study individual ageing over the lifespan of a single animal; however, it does not constitute a robust cross-sectional population biomarker (Dunshea *et al.* 2011; Horn *et al.* 2008). This is due to variation in telomere length at young ages and external effects influencing individual telomere attrition differently (Haussmann & Mauck 2008). Sex and variable age-specific patterns of telomere loss have been reported in the thick-billed murre (*Uria lomvia*) (Young *et al.* 2013) and wandering albatross (*Diomedea exulans*) (Hall *et al.* 2004). Several molecular methods of ageing were reviewed by Jarman et al. (Jarman *et al.* 2015).

Several recent studies demonstrate that specific epigenetic modifications can be used to determine individual chronological age (Horvath 2013; Maegawa *et al.* 2010). DNA methylation (DNAm) is the most studied epigenetic mark and involves the addition of a methyl group to the 5' cytosine (5mC), of cytosine – guanine dinucleotides, hereafter referred to as CpGs (Jones *et al.* 2015). The presence or absence of DNAm in important gene regions such as promoters can lead to changes in gene expression. Generally, higher levels of DNAm (hypermethylation) of promoter CpG islands results in stable gene suppression (Jones 2012). Methylation changes at specific CpGs have previously been linked to chronological age in humans (Grönniger *et al.* 2010; Hannum *et al.* 2013; Horvath 2013), mice (Maegawa *et al.* 2010), humpback whales (Polanowski *et al.* 2014), dogs (Jakubczak *et al.* 2016) and developmental stage of the bat wing (Eckalbar *et al.* 2016). Identification of DNAm changes in tissues that can be sampled non-lethally such as skin (Grönniger *et al.* 2010; Polanowski *et al.* 2014) and blood (Bekaert *et al.* 2015; Weidner *et al.* 2014) provides the opportunity to develop an epigenetic model to estimate age in live seabirds.

The short-tailed shearwater (*Ardenna tenuirostris*) (Christidis & Boles 2008; Penhallurick & Wink 2004), is the most abundant seabird species in Australian waters and are one of the few native Australian birds still harvested (Skira *et al.* 1996). While there are many useful methods for age estimation in wild animals some are only applicable to dead animals or have a large margin of error (Campana 2001; Jarman *et al.* 2015). Some birds exhibit external features that allow the estimation of age or developmental stage (Møller & De Lope 1999), however the majority of seabirds, including shearwaters, have no morphological marks for age determination. Age estimation methods for colonial seabirds therefore rely on extensive and often labour intensive long-term banding studies (Bradley *et al.* 1991). Shearwaters breed at approximately five years of age in Australia and each year migrate 15,000 km to coastal areas around the Aleutian Islands and Kamchatka Peninsula (Marshall & Serventy 1956). Adults and newly fledged birds typically return to the same areas to breed if they survive the long migration; therefore, individuals can be recaptured allowing for both longitudinal and cross-sectional sampling (Bradley *et al.* 1991). These features make the shearwater an ideal species to test the applicability of mammalian age-related methylation patterns in birds.

We investigated the methylation levels of CpGs in several genes using DNA isolated from shearwater blood and feather tips. By using a closely related seabird genome, we identified a number of CpG sites within a set of candidate genes known to have age-related changes in DNAm in mammals. The DNAm levels observed in the long-lived seabird *A. tenuirostris*, did not support the hypothesis that DNAm changes with age recorded in mammals were conserved between the two vertebrate classes for these genes.

3.3 Materials and methods

3.3.1 Study sites

Blood and feather samples were collected from short-tailed shearwaters on Fisher Island, Tasmania (40°13'00.8"S 148°14'21.1"E) and from Fort Direction, Tasmania (43°02'48.7"S 147°24'58.5"E) under Department of Primary Industries, Parks, Water and Environment (DPIPWE) permit: FA 15230 and University of Tasmania (UTAS) Animal Ethics Committee permit: A14277. Samples for this study were collected from adults during incubation in December 2014, 2015 and from chicks prior to fledging in March 2016. During the surveys, all burrows were inspected by hand to determine occupancy, then the weight of the occupant and leg band number, if present, were recorded. During handling, birds were restrained in a dark, air-permeable bag and returned physically to their burrow of origin.

3.3.2 Sampling

All individuals were weighed using a calibrated 1000 g or 2500 g Pesola handspring scale; blood samples were only taken from animals with a weight greater than 500 g. Venepuncture of the vein in the webbing of the foot was done with a 25-gauge needle. A few drops of whole blood (approximately 0.1 – 0.2 mL) were collected on a Whatman FTA® Micro (WB120210) card, dried and stored in sealable plastic bags. Three to five breast feathers were plucked and preserved in salt saturated dimethyl sulfoxide (DESS) solution. 31 known age feather samples were collected from Fisher Island in 2014 and 30 matched feather and blood samples from known age birds were collected in 2015. A further 20 matched chick blood and feather samples were also collected in 2016 from separate burrows at Fort Direction. The sex of the bird associated with each sample was determined with CHD1 real-time PCR and melt curve analysis (Faux *et al.* 2014).

3.3.3 DNA extraction

Blood samples immobilised on FTA cards were extracted using Epicentre QuickExtract™ (QE09050) or MasterPure™ (MCD85201) DNA Purification Kits. For MasterPure™ extractions, a 3 mm punch was cut from an FTA card using a sterile punch and placed into a lysis solution containing Proteinase K and DNA was isolated according the manufacture's instructions. DNA was eluted in 20 µL of TE. This protocol was also used for the extraction of 1 – 3 feather quill tips. Approximately 1 – 2 cm of the quill end was dissected with a sterile scalpel and placed into digestion buffer. Feather samples were centrifuged for an additional 5 minutes at 10,000 g at room temperature following protein precipitation and supernatant isolation to ensure no debris were carried through. 1 – 2 µL of isolated DNA was quantified with high sensitivity or broad range reagent kits for the Qubit® 2.0 Fluorometer and standardised to 50 ng/µL with TE.

Isolated feather or blood DNA (50 ng - 1 µg) was bisulphite converted for DNA methylation analysis using a Zymo EZ DNA Methylation-Lightning™ Kit (D5030) following the manufacturers' instructions. A dilution experiment was prepared using feather DNA for this protocol. Isolated genomic DNA was serially diluted in TE buffer at the following dilutions; 1:1, 1:5, 1:10, 1:100 and 1:1000 and subsequently bisulphite converted and amplified using shearwater bisulphite specific primers.

3.3.4 Primer design

Candidate genes were identified from previously published data supporting an association between DNAm and age in mammals for specific CpG loci. Data published using the Illumina HumanMethylation 450k or 27k Array allowed for the ready identification of the CpG loci under investigation using the Cluster CG number and associated sequence. For other loci, position numbers relative to the transcription start site were used to identify their location within the gene. Target CpG conservation was checked using the University of California Santa Cruz (UCSC) Genome Browser (Kent *et al.* 2002) for the annotated human assembly (hg38) or chicken (*Gallus gallus*) assembly (galGal4). An unannotated genome (Assembly: ASM69083v1), sequenced to a depth of 33x, is available for the northern fulmar (*Fulmarus glacialis*), a long-lived seabird also of the Procellariidae family. As the shearwater at the time of writing did not have a published reference genome, sequences from the

northern fulmar were used to design primers to amplify shearwater gene regions. If a target CpG was conserved in the orthologous chicken or fulmar gene the Reference Sequence (RefSeq) (Pruitt *et al.* 2012) was used as input for the online BLASTn suite (Altschul *et al.* 1990). Highly conserved sequences were returned from BLAST using the inbuilt Gnomon gene prediction tool to facilitate identification of orthologous genes (Souvorov *et al.* 2010). These sequences were then checked for CpG loci, including any previously published loci and used to design primers for shearwater genomic DNA amplification.

Primers designed from northern fulmar reference sequences shown in Table A in S1 File were created using the online client of Primer3Plus (vsn.2.4.0) (Untergasser *et al.* 2012) and were split into *gene set 1* and *2*. *Gene set 1* consisted of initial mammalian gene targets, whereas *gene set 2* consisted of additional amplicons for ELOVL2, TET2 and two genes new to this study, MYOD1 and TRIM59. CpG loci were specifically excluded from primer pairs at this stage to simplify downstream design of bisulphite-converted primers. Each primer pair designed using this method was optimised with a temperature gradient (56.1°C – 64.7°C), no template control (NTC) and sequenced on an Applied Biosystems Genetic Analyzer 3130. Sequences were visualised in Sequencher (vsn. 4.10.1) and BioEdit (vsn. 7.2.5) (Hall 1999) to identify base pair differences between the two birds and a short reference sequence for the shearwater was created. Once genomic amplification was achieved the new shearwater sequence was used as a reference to create new primers in MethPrimer (Li & Dahiya 2002) and Zymo Bisulfite Primer Seeker specific to bisulphite converted DNA (Table B in S1 File). Bisulphite specific primers were temperature optimised (50.2°C – 58.7°C) and run with NTC and genomic DNA controls to ensure specificity. All primer optimisation was visualised using pre-stain electrophoresis on 1.5% agarose gels running at 100 V for 50 minutes with SYBR® Safe (Thermo Fisher Scientific) or GelGreen™ (Biotium) dyes. PCR cycling conditions for genomic, bisulphite converted DNA and library preparation are described in Tables C-E in S1 File.

3.3.5 Library preparation

To prepare the avian samples for targeted bisulphite sequencing on the Illumina MiSeq Platform, genomic DNA was bisulphite converted as described above. Up to 50 ng and 500 ng of feather and blood genomic DNA respectively was converted and eluted in 10 uL of M-

elution buffer. Post-PCR products were diluted 1:10 and in-house identifier sequences were added to the end of the MiSeq universal primers with a ten round PCR. Samples were then pooled into a single 1.5 mL Eppendorf tube and diluted to 2 nmol/L. The library was prepared for sequencing on a Nano, Micro or Standard flow cell (Illumina) following the manufacturer's instructions with 20% PhiX control.

3.3.6 Data analysis

Fastq files were unpacked and Phred (Q) scores were assessed using commands for USEARCH v8.1 (Edgar 2010). A stringent maximum expected error of 0.7 was used to filter poor quality reads. Known CpG sites were identified within target amplicons using a second script that used a unique 'core' sequence approximately five base pairs long before the CpG site. This approach reduces the chance of losing informative reads due to read end differences or errors in large base pair repeat areas. The methylation level for each site was calculated as number of methylated cytosines divided by the total number of reads for the given CpG pair and recorded between 0 and 1, where 0 is unmethylated and 1 is methylated (S10 File). A read depth of >100 was required to include a score in the final analysis; scores with a read count below this were discarded.

Linear regression of age and DNAm levels for CpG loci in amplicons from both gene sets was executed in R to aid in visualising the data. The R package 'glmnet' was then used to fit penalised lasso regularisation paths to variables in a generalised linear model (S11 File) (Friedman *et al.* 2010). We used a matrix containing all calculated ratios (predictors) for every CpG site for each gene target. Glmnet was then used to fit the DNAm data to age using alpha = 1 (lasso). In order to select λ (lambda), the penalty value, the cross validation function of glmnet was used. A penalised statistical method was used as the number of CpG sites scored was greater than the number of individuals used in the study.

3.4 Results

3.4.1 DNA yield from shearwater blood and feather samples

A total of 53 feather (47 adult, 6 chick) and 32 blood (30 adult, 2 chick) samples from individuals of known age were used to investigate the DNA methylation status of target

genes. CpG loci in *gene set 1* (see Table B in S1 File) were investigated using DNA from 36 feather samples including 6 chicks (40 – 50 days old) and 30 adults (mean age = 11.7 years, range: 5 – 21). Matched blood samples from 14 birds were analysed in the same gene set (adult mean age = 12.5 years). Targets in *gene set 2* were analysed using DNA from 42 adult (mean age = 11.3 years, range: 2 – 21) and 2 chick feather samples. Blood samples from 32 individuals (adult mean age = 11.7 years, range 5 – 21) were also analysed for this gene set, 27 of these were matched to feather samples (Table 1). The sex of 27/30 (90%) known-age adults was determined using a real time PCR assay from blood spots (Faux *et al.* 2014). S1 Fig shows an example of a successful melt curve analysis following real time amplification.

Table 1. Feather and blood samples used.

	Feather		**Blood**		**Matched**	
Gene set	Chick	Adult	Chick	Adult	Chick	Adult
1	6	30	2	12	2	12
2	2	42	2	30	2	25

DNA yield was greater in feathers collected from chicks with an average total yield of 1844 ng (n = 6, ± 343) from one feather compared to an average of 124 ng (n = 47, ± 117) from adult feathers using the same extraction method (Fig 1A, Table F in S1 File, t-test: $p < 0.0001$). Total DNA yield was consistently high from shearwater blood samples stored on FTA cards with an average of 1826 ng (n = 30, ± 1295, Fig 1A, Table F in S1 File). DNA yield from adult feathers was dependent on the number of quill tips prepared for digestion. Extraction of a single quill tip yielded an average of 68 ng (n = 23, ± 34) of DNA whilst a two and three quill tip extraction yielded 149 ng (n = 25, ± 141) and 179 ng (n = 9, ± 70) of DNA respectively (Fig 1B, Table G in S1 File). Yield from one feather was compared to two (ANOVA: $p = 0.0197$) and three (ANOVA: $p = 0.0194$) quill tips. S2 Fig shows a quality comparison of DNA isolated from different avian tissues. Input of approximately 50 ng of

DNA showed a high molecular weight band of genomic DNA when run on a 1% agarose gel indicating the successful isolation of high quality DNA.

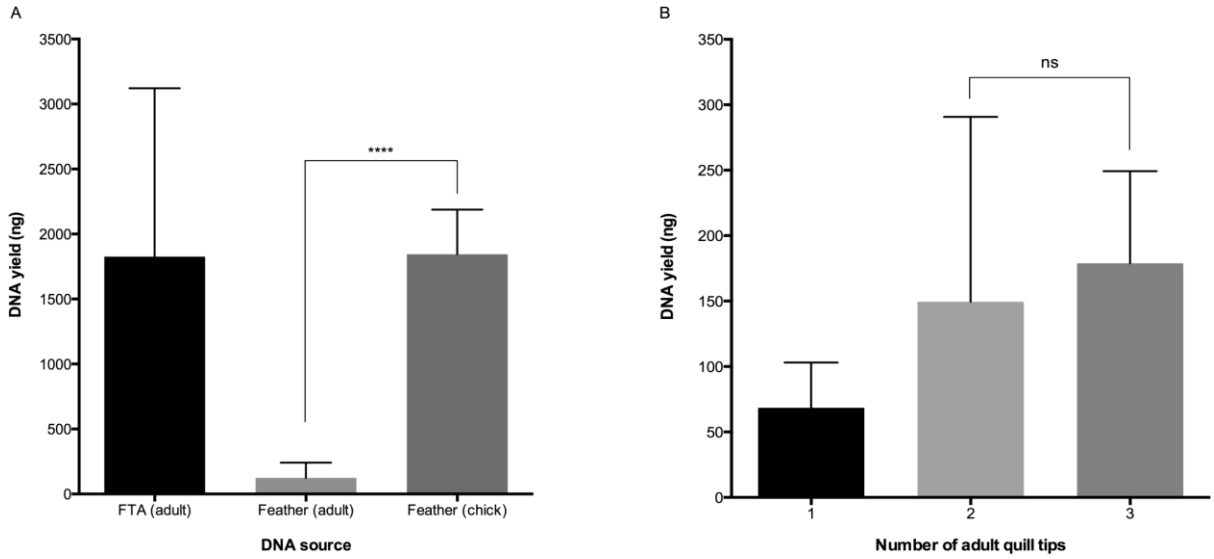


Figure 1 | DNA yield varies with tissue source and age. A: Genomic DNA was isolated from blood stored on FTA cards (n = 132) and feather quill tips adult (n = 58) and chick (n = 6). DNA yield from adult feathers was significantly lower than chicks. **B**: The isolation of sufficient genomic material for experimental optimisation, bisulphite treatment and methylation analysis increases with the number of adult feather quill tips included in the extraction solution.

3.4.2 Amplification of shearwater DNA

For *gene set 1* a total of 13 genomic primers were designed from the fulmar sequence to target the first exon and promoter regions of each gene. Eight of these primers amplified the target whilst five failed to amplify template DNA. In *gene set 2*, seven primers were designed with two failing to amplify the correct product. In both gene sets positive products were sequenced and verified by BLASTn against the northern fulmar or chicken sequence. BLASTn metrics indicating the degree of conservation between fulmar/shearwater and human genomic regions are shown in Table K in S1 File.

Following amplification of target genomic DNA the products were sequenced on an ABI3130. This generated several short reference sequences from which we could design primers specific to bisulphite converted DNA. In *gene set 1*, nine primers were designed from the reference sequences and all but one of these amplified the correct converted product. In *gene set 2*, a further six primers were made and again one failed to amplify the correct product.

Overall there was relatively poor sequence conservation between mammalian and avian CpG sites. An example of sequence conservation between human, mouse, chicken, fulmar and shearwater sequences is shown for MYOD1 in Fig 2A. This figure shows the age associated cg18555440 and a further six CpG sites within the MYOD1 gene. Five of the sites are conserved within the chicken genome but the northern fulmar and shearwater sequence show further differences with only three sites of the original six present in shearwaters.

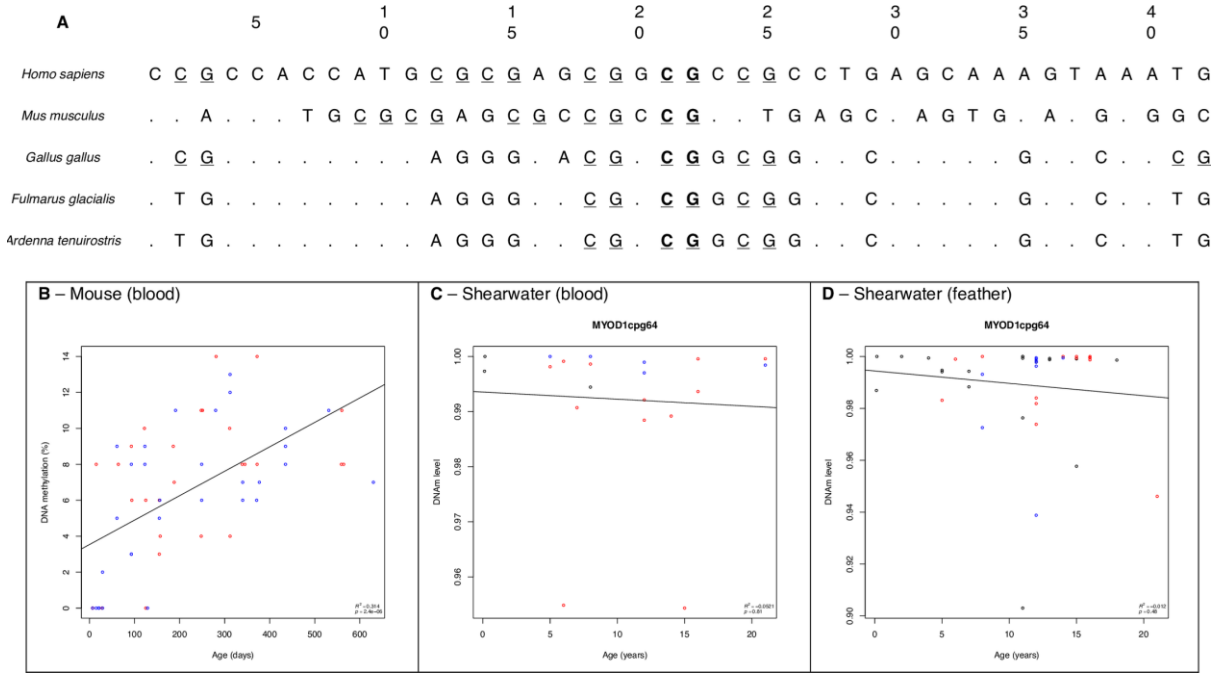


Figure 2 | *MYOD1* sequence conservation in mammals and select birds. A: 20 base pairs are shown for both directions around an age related CpG loci (cg18555440, in bold) in the *MYOD1* gene in humans. Following the methods described in text, this sequence was compared to the mouse, chicken and northern fulmar genomes. Conserved bases are shown with a dot (.) and the base is given where there is a mismatch. Other CpG sites in the sequence are underlined and are shown for bird sequences. In this example, 4 CpG loci in humans were conserved in chicken, 2 were lost (pos. 12 and 14) and 1 was gained (pos. 41). However in fulmar and shearwater, 3 CpG loci were conserved whilst the remaining 3 in humans were lost (pos. 2, 12 and 14). Primers were designed from the fulmar sequence to amplify isolated shearwater DNA; for this short sequence there was 100% conservation between the two animals. A plot of the age relationship observed in mice at cg18555440 (Spiers et al. 2016) **Fig 2B**. The DNAm results for the same CpG site in shearwater are shown for **shearwater blood, Fig 2C** and **shearwater feather, Fig 2D**. Males and females are shown in red and blue respectively.

3.4.3 Analysis of DNA methylation and age

As very little feather DNA was available for bisulphite treatment, which is known to be particularly harsh on DNA, we prepared a serial dilution of bisulphite-converted feather DNA to assess the potential amplification quality of diluted sample. Dilutions of a 10 ng/µL bisulphite converted DNA sample at 1:5, 1:10, 1:100 and 1:1000 all showed the correct 243 bp product for ASPA (S3 Fig). Based upon this result, subsequent bisulphite PCR used a 1:5 dilution of converted DNA (approximately 2 ng).

For *gene set 1* blood and feather, run on Standard and Micro flow cells, the range of reads was approximately 10,000 – 100,000 and 500 – 10,000 reads respectively. For *gene set 2* all samples were run on the same micro flow cell and a standardised PCR input led to a read depth of 500 – 2000 reads (average reads per tissue shown in Fig 3). Two replicate feather samples were sequenced in *gene set 1*, a technical replicate (same individual different feather quill tip) for a chick and a run replicate (same DNA extraction) for a 12-year-old adult. The absolute difference in the reported DNAm level was calculated for each replicate as shown in Table H in S1 File. We report a mean difference of 5.84% and 5.04% in DNAm levels for the technical and run repeat respectively.

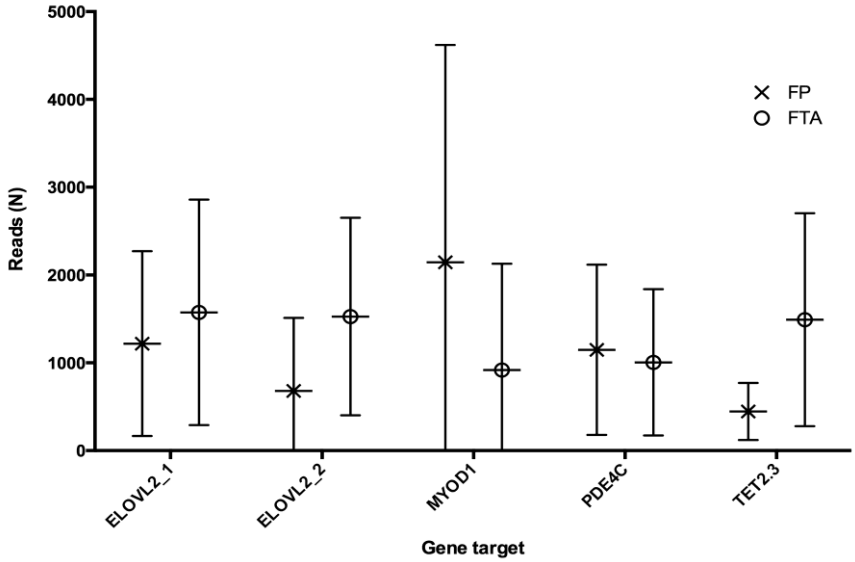


Figure 3 | Total reads used to calculate methylation level in *gene set 2*. The average number of reads varied with each amplicon and tissue type (FP: feather, FTA: whole blood) being investigated. If the total number of reads for a single CpG loci was <100 the result was removed from further analysis.

The relationship (R^2 = 0.314, p = 2.4e-06) between age and DNA methylation in an inbred mouse model for cg18555440 (Spiers *et al.* 2016) is shown in Fig 2B. This result was directly compared with DNAm levels in blood and feather in our study (Fig 2C-D), indicating that an age relationship at this site is not conserved in *A. tenuirostris*. Fig 4A shows a similar comparison for the *KCNC3* gene and shows the loss of three CpG sites in fulmar and shearwater sequences. DNAm levels are shown for cg06572160 in humans (Fig 4B) as this site was previously associated with age and conserved in shearwater (Fig 4C-D, CpG66) (Koch & Wagner 2011; Rakyan *et al.* 2010). Sequences for ASPA were also compared in Panels A-C in S4 Fig.

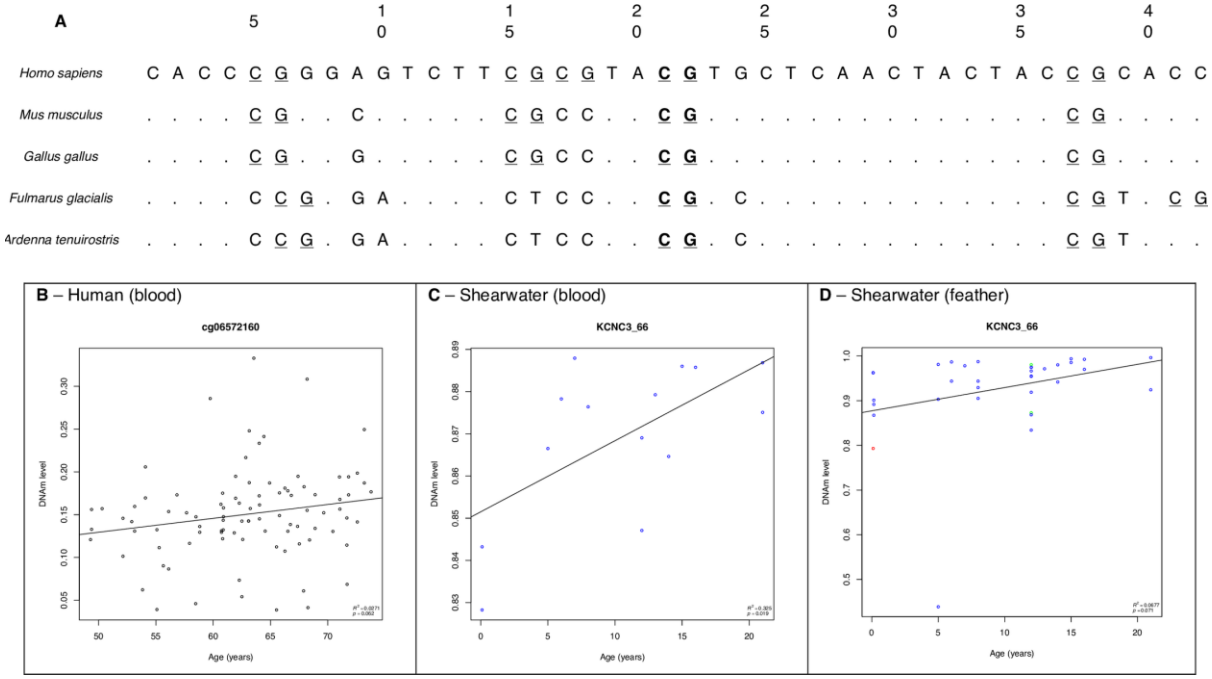


Figure 4 | *KCNC3* sequence conservation in mammals and select bird species. A: 20 base pairs are shown for 3' and 5' directions around an age related CpG loci (27K: cg06572160, in bold) in the *KCNC3* gene in humans (*KCNC4/2* in *F. glacialis* and *G. gallus*). Following the methods described in text, this sequence was compared to the mouse, chicken and northern fulmar genomes. Conserved bases are shown with a dot (.) and the base is given where there is a mismatch. Other CpG sites in the sequence are underlined and are shown in the bird sequences. In this example, 4 CpG loci in humans are conserved in chicken, whilst 1 is lost (pos. 17). However, in fulmar and shearwater, 3 CpG loci were conserved whilst the remaining 2 in humans were lost (pos. 15 and 17). A plot of the age relationship observed in **human blood, Fig 4B** at cg06572160 (Rakyan et al. 2010). DNAm results for the same CpG site in shearwater is shown for **shearwater blood, Fig 4C** and **shearwater feather, Fig 4D**. A technical replicate is shown in red whilst a run replicate is shown in green for **D**.

The age distribution of samples used for the following results is shown in S5 Fig. The top nine CpG sites demonstrating the strongest relationship with age are presented in Fig 5. The highest coefficient of determination (R^2) was 0.325 (p = 0.019) in CpG66 from *KCNC3* amplified from whole blood (DNAm range: 0.83 – 0.89). Five of the top 9 CpG loci were identified in *KCNC3* blood samples. CpG42 in *ELOVL2* (*gene set 2*) feather tissue was the next best site with R^2 = 0.285 (p = 0.00048) but interestingly, we found a greater magnitude of DNAm change with age (range = 0.79). Three CpG loci from the same *ELOVL2* amplicon, in both feather and blood samples, appeared in the top results and all showed a conservation of the broad range of DNAm levels observed in CpG42. CpG46 in *EDARADD* feather tissue had an R^2 = 0.168 (p = 0.0067) and a DNAm range of 0.85 – 1.0. Regression plots for *gene set 1* and *gene set 2* blood and feather CpG loci can be viewed in S2 – S5 File respectively along with the collated R^2 values in Tables I-J in S1 File and raw data in S6 – S9 File respectively. A penalised lasso model was fitted to the collected data; however, due to the relatively poor correlations with age no CpG sites were selected at minimum or 1se values of λ (factor selection cut-off).

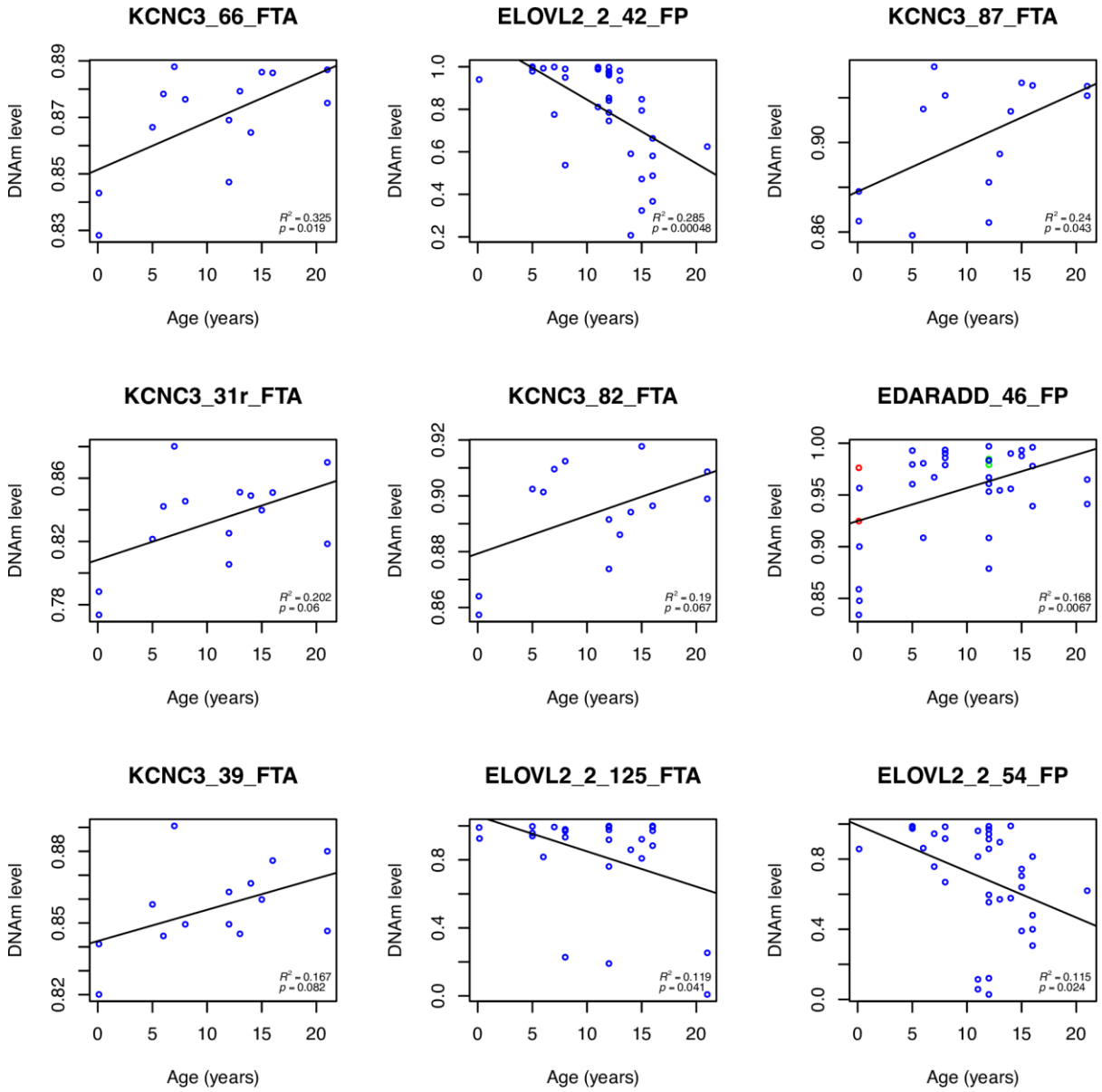


Figure 5 | Methylation profiles by MiSeq analysis. Association of individual seabird proportion of cytosine methylation (*y*-axis) against age in years (*x*-axis) for the top nine CpG sites in blood (FTA) and feather (FP) samples. The best relationship with age based on R^2 is shown from the top left to bottom right. A measure of the magnitude of DNAm change with age is also shown. A technical replicate is shown in red whilst a run replicate is shown in green in the *EDARADD* methylation plot.

Comparisons of average DNAm levels for highly reported loci were made between birds and humans where possible. Human cg02228185 in ASPA was compared to three nearby sites for the shearwater indicating large DNAm differences (Fig 6A). A direct CpG site DNAm comparison is shown between human and shearwater in Fig 6B, again showing a large difference in methylation level. A number of human promoter CpGs in *ELOVL2*, a frequently reported gene associated with age, were also compared to several downstream sites in shearwater (Fig 6C). A higher level of variation in DNAm was observed for these sites.

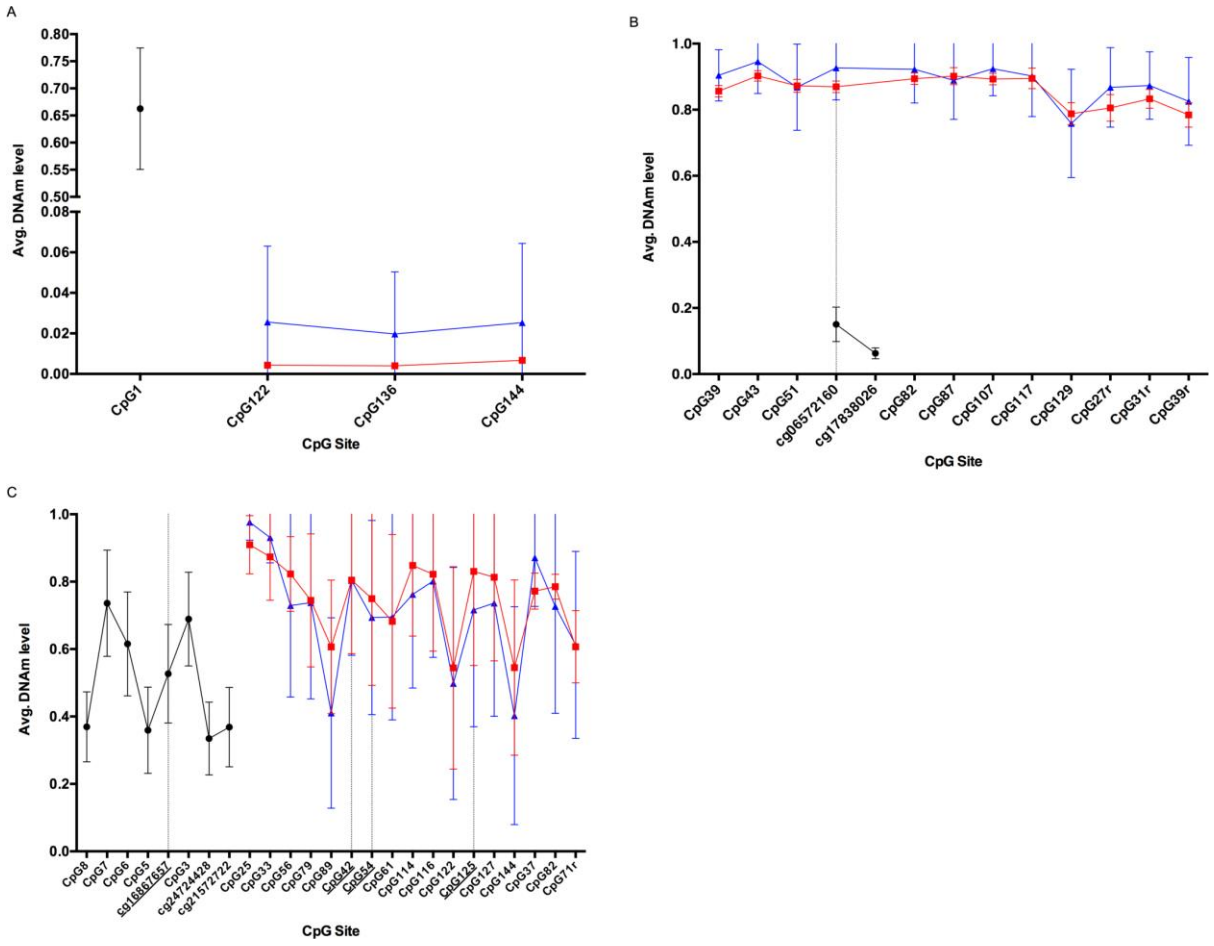


Figure 6 | CpG site average methylation level comparisons. Human samples are shown in black, whilst shearwater blood and feather in red and blue respectively. **A**: Comparison of cg02228185 (CpG1) in *ASPA* analysed in Bekaert et al. (2015) and downstream CpG sites analysed in shearwater shows large differences in methylation levels. CpG1 was not conserved in the shearwater sequence (C>A in chicken, fulmar and shearwater). **B**: A direct comparison of cg06572160 in *KCNC3* analysed in Rakyan et al. (2010). Average DNAm levels in shearwater blood and feather samples were higher than in human blood. **C**: Average DNAm levels for a number of highly reported age related CpG sites are shown for *ELOVL2* in human blood (Bekaert et al. 2015). Whilst these markers could not be directly compared with the avian equivalent, CpG sites are shown downstream for both shearwater tissues. Underlined CpGs indicate those that are highly correlated with age in humans and appeared in the top nine results in our study.

DNAm levels were also compared between the two tissue types analysed, feather and blood. We observed large variances in DNAm levels between different genes as represented by randomly selected CpG loci within those genes (Fig 7A-C). We observed relatively low and consistent levels of methylation in both feather and blood tissue for *ASPA* (CpG122 range: 0 – 0.15). In contrast, relatively high and stable levels of methylation were observed in *TET2* amplicons as represented by CpG57 (range: 0.5 – 1.0). CpGs in *ELOVL2* showed increased deviation in DNAm levels (range 0.2 – 1.0) and decreased parsimony between the tissues compared to loci in TET2 and ASPA.

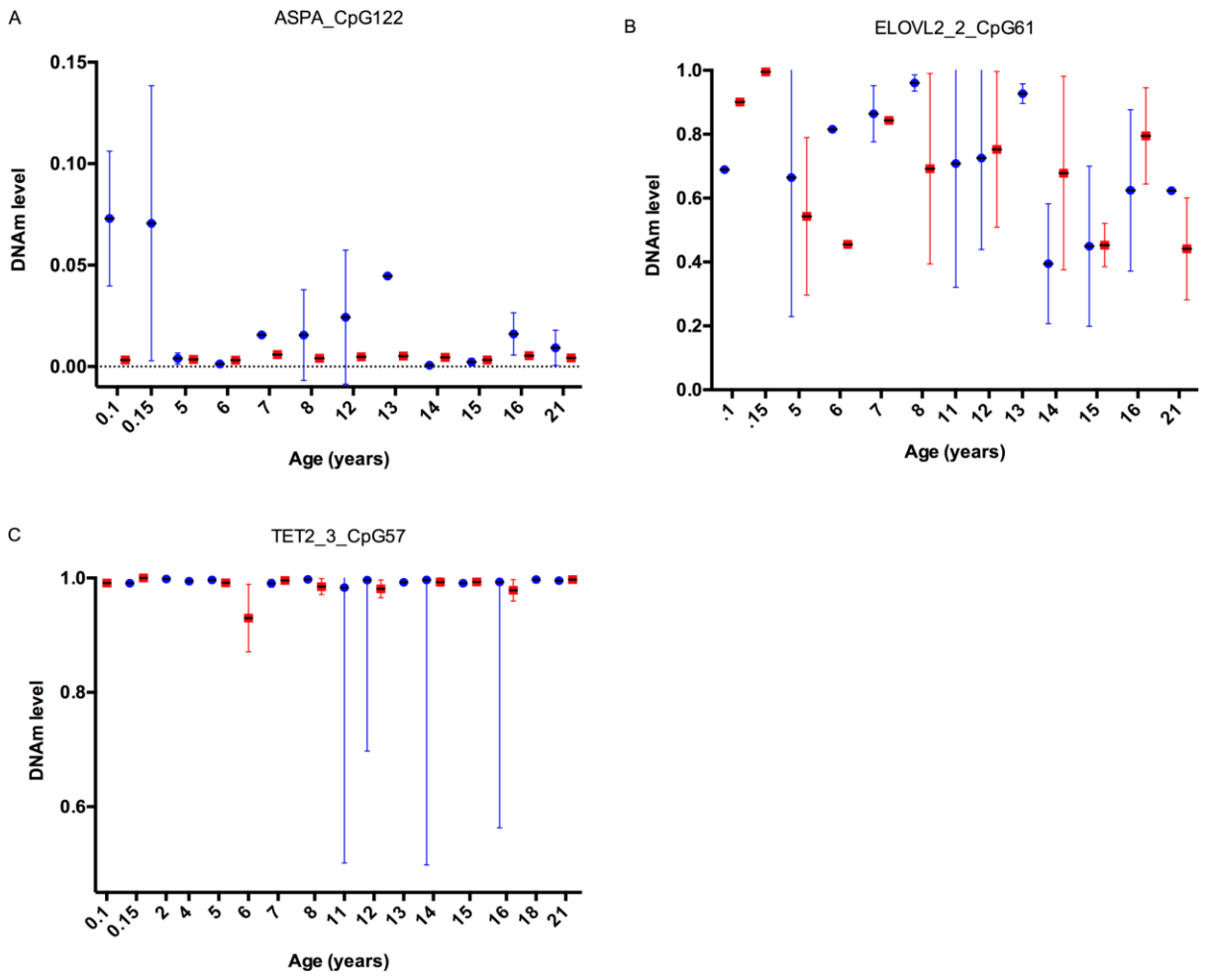


Figure 7 | Comparison of DNA methylation levels in different avian tissues. We observed a large range of amplicon dependent DNA methylation levels. Here, the average DNAm levels for each age analysed in feather (blue) and whole blood (red) were plotted side by side to assess how conserved these signals are. Relatively low, stable (between tissue types) methylation levels are shown in **A)** for CpG 122 in *ASPA* from *gene set 1*, variable and less stable methylation levels are shown in **B)** for CpG 61 in the second *ELOVL2* amplicon from *gene set 2* and finally relatively high, stable methylation levels are shown in **C)** for CpG 57 in the third *TET2* amplicon in *gene set 2*.

3.5 Discussion

This study is the first, to our knowledge, to target known mammalian age-related genes in a wild bird population. We successfully analysed methylation levels in 67 CpGs in 13 genes, but despite finding weak correlation with age in some sites, we were unable to create a model to estimate age using CpG methylation as done in other studies (Polanowski *et al.* 2014; Weidner *et al.* 2014; Zbieć-Piekarska *et al.* 2015b). Age has been associated with DNA methylation levels in a number of different mammalian tissues including blood (Bekaert *et al.* 2015; Garagnani *et al.* 2012; Hannum *et al.* 2013; Horvath 2013; Zbieć-Piekarska *et al.* 2015b), skin (Grönniger *et al.* 2010; Polanowski *et al.* 2014), saliva (Bocklandt *et al.* 2011) and dentin (Bekaert *et al.* 2015); we sought to investigate DNAm levels from two avian tissues: blood and feathers.

Blood samples provide a valuable source of DNA as avian red blood cells are nucleated and can yield up to 1.8 μg per 3 mm punch of an FTA card; greater than the DNA yield per equal volume of mammalian blood (Mas *et al.* 2007). Current field techniques for blood collection generally require controlled low temperature storage in difficult climates or in locations with limited or no electricity. Our results further support the use of FTA storage cards that immobilise blood and pathogens on a paper-like matrix (Smith & Burgoyne 2004).

Feathers are a promising DNA source as they can provide DNA from the skin and recent studies have also shown that DNAm in mammalian skin samples are strongly correlated with age (Grönniger *et al.* 2010; Hannum *et al.* 2013; Koch & Wagner 2011; Polanowski *et al.* 2014). While feathers are generally considered to be a more minimally invasive technique than taking blood samples they do present their own problems (McDonald & Griffith 2011; Smith *et al.* 2003). First, our study shows the relatively low yield of high quality DNA can be limiting for repeated applications such as bisulphite sequencing. Second, condition of the feather and length of the sequence has been reported to affect PCR success (Beja-Pereira *et al.* 2009), with one study reporting 60% and 50% amplification success with plucked and moulted feather respectively (Segelbacher 2002). Of the failed PCRs in this study, sequence length did not appear to play a role. Third, low yield and quality, can lead to unnecessary laboratory repetition and ultimately re-sampling of bird populations that otherwise may not be required (McDonald & Griffith 2011). In our study, DNA yield from feathers was an issue

as there would generally not be enough material for more than two sequencing runs. Given that our results indicate some differences between blood and feather, we would suggest that multiple tissues continue to be tested until a biomarker of age is discovered. But the feasibility of feathers as a tissue source for ageing needs to be assessed for each study. Further, the blood sampling method described allows the long-term storage of samples and numerous re-extraction opportunities from a single blood spot allowing access for future techniques or validation studies (Smith & Burgoyne 2004).

Amplifying age-related regions in birds based on previously defined mammalian targets can be challenging due to poor sequence conservation. Indeed, the examples shown in Figs 2 and 3 were the only cases of exact CpG site conservation in our bird sequences. Our study was helped by the availability of a closely related bird genome sequence. These examples highlight the difficulties faced when targeting specific mammalian biomarkers of age in non-mammals, but also shows that these targets can provide a solid foundation from which to begin investigations. Previously, primers used to amplify conserved sequences in the humpback whale (*Megaptera novaeangliae*) were designed using a closely related dolphin species (*Tursiops truncatus*) (Polanowski *et al.* 2014). In this study, samples were then analysed with the PyroMark system, which compared to the MiSeq can have a higher read length but lower throughput (Liu *et al.* 2012). Using the Illumina MiSeq Micro and Nano flow cells we were able to sequence to an average depth of 2100 reads per sample. This technique allowed us to calculate DNAm in a similar way to other studies (Zbieć-Piekarska *et al.* 2015b).

In shearwater blood five of the top nine age-related loci came from CpG sites in the *KCNC3* locus. This gene has previously been identified to be age related (cg06572160, $r = 0.70$) in human dermis, epidermis and blood (Koch & Wagner 2011). The protein encoded by *KCNC3* is an integral membrane protein that mediates potassium ion permeability and has an established role in adult-onset neurodegeneration in humans (Waters *et al.* 2006). Hypermethylation of this gene may support its role in the gradual decline of cellular function with age (Koch & Wagner 2011). It is interesting to note that whilst we observed changing DNAm, the overall level for this site is much higher in *A. tenuirostris* than in humans, possibly indicating regulation differences. In shearwaters the DNAm range (0.78 – 0.94) is

quite small, so methylation levels need to be measured with large sequencing depth to detect differences. Additionally, samples from chicks are included and influence the slope whilst adults do not show a strong relationship (e.g. CpGs 87 and 82), and therefore offers limited predictive capacity. Some CpG sites (e.g. 31r and 39) showed a stronger hypermethylation trend in adults without inclusion of chick data.

The greatest difference in methylation levels with age (range: 0.2 – 1) were found in CpG sites in the fatty acid elongase 2 gene (*ELOVL2*) and were the second most reported in the top nine loci. Hypermethylation of sites in this gene during aging has been reported in human blood (cg16867657, $r = 0.91$) (Garagnani *et al.* 2012) and is being considered for forensic application (Zbieć-Piekarska *et al.* 2015a). *ELOVL2* and *5* have been characterised in chicken liver and reported to be different from other species due to their ability to elongate docosapentaenoic acid, however no methylation studies have been carried out (Gregory *et al.* 2013). Our results contrast with previous mammalian studies as the strongest observed change was observed in DNA isolated from quill tips (dermis) rather than blood and shows hypomethylation with age. Therefore, whether methylation is an indication of age only or is functionally correlated with ageing requires further study. Another top gene identified was EDAR associated death domain (*EDARADD*). CpG sites from *EDARADD* have been previously identified to correlate with age in human blood (cg09809672) (Vidal-Bralo *et al.* 2016) and saliva samples ($r = -0.81$) (Bocklandt *et al.* 2011). However, in our study we observed DNAm saturation for older individuals and relatively large variation in chick samples indicating that it is unlikely to be of use for shearwater ageing.

Shearwaters have a maximum lifespan of 40 years. Our samples covered approximately half of this range (chicks – 21 years). Previous models report age estimates with an error of approximately 3 – 6% the lifespan of the animal (Hannum *et al.* 2013; Polanowski *et al.* 2014). If a strong age relationship were present in the sites investigated it would be expected to estimate age in birds that fall within this range. Nonetheless, our study could be strengthened by the addition of older individuals. Unfortunately the chance of capturing older individuals diminishes with each season due to mortality (Bradley *et al.* 1991). Despite the shorter lifespan of the shearwater compared to humans and whales, ageing-related DNAm changes are theoretically detectable. Several genes have been identified in mice which have

DNAm changes on specific cytosines from a number of different tissues from mice ranging from 2.9 to 35.2 months (Maegawa *et al.* 2010). This evidence would suggest that the DNA methylation changes observed in ageing are conserved between short and long-lived animals (Jarman *et al.* 2015).

In this study, we focused on DNAm levels at specific sites as this approach has been used for all molecular age biomarkers developed so far (Koch & Wagner 2011; Polanowski *et al.* 2014; Weidner *et al.* 2014). However, an alternative approach is to investigate genome-wide DNA methylation, which in two studies of chickens (*Gallus gallus*) has indicated a relationship with age (Gryzińska *et al.* 2013; Li *et al.* 2011). Whole-genome scale methylation observations from 7-day-old chickens indicated that chicken tissues generally displayed methylation patterns similar to that in other animals. Compared to the pattern in mammals, DNAm was enriched in gene bodies and promoter methylation was negatively correlated with gene expression, indicating a suppressive role (Li *et al.* 2011). Genome-wide studies have so far focused on investigating developmental or tissue specific methylation and have included few adult samples. An observation from one study showed differing overall methylation levels between tissues in three strains of 14-week-old chickens (Xu *et al.* 2007). These studies further our understanding of bird epigenetics and indicate that genome-wide techniques are a suitable approach for the detection of DNA methylation and therefore may be suitable for discovery of age-related genes.

In conclusion, we present a widely applicable and high-throughput technique to quantify DNAm levels in candidate age-related genes in long-lived seabirds. While we were unable to identify a combination of CpGs that would constitute a suitably strong biomarker of age in the short-tailed shearwater, we have identified several CpGs that were weakly correlated with age that warrant follow up investigation with a larger number of samples or wider age range. Our results show that an approach focusing on candidate genes identified in mammals is not a straightforward way to identify genes with age-related DNAm changes in birds. Future studies of a greater number of CpG pairs should expand the list of age-related sites in seabirds. Alternative methylation analysis techniques such as digital restriction enzyme methylation analysis (DREAM) and reduced representation bisulphite sequencing (RRBS) could be used.

Chapter 4 - Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers

Published as: R. De Paoli-Iseppi, B.E. Deagle, A.M. Polanowski, C.R. McMahon, J.L. Dickinson, M.A. Hindell, S.N. Jarman. (21 December 2018). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular Ecology Resources*. (In press).

4.1 Abstract

Age structure is a fundamental aspect of animal population biology. Age is strongly related to individual physiological condition, reproductive potential and mortality rate. Currently, there are no robust molecular methods for age estimation in birds. Instead, individuals must be ringed as chicks to establish known-age populations, which is a labour intensive and expensive process. The estimation of chronological age using DNA methylation is emerging as a robust approach in mammals including humans, mice and some non-model species. Here we quantified DNA methylation in whole blood samples from a total of 71 known-age Short-tailed shearwaters (*Ardenna tenuirostris*) using digital restriction enzyme analysis of methylation (DREAM). The DREAM method measures DNA methylation levels at thousands of CpG dinucleotides throughout the genome. We identified seven CpG sites with DNA methylation levels that correlated with age. A model based on these relationships estimated age with a mean difference of 2.8 years to known age, based on validation estimates from models created by repeated sampling of training and validation data subsets. Longitudinal observation of individuals re-sampled over 1 or 2 years generally showed an increase in estimated age (6/7 cases). For the first time, we have shown that epigenetic changes with age can be detected in a wild bird. This approach should be of broad interest to researchers studying age biomarkers in non-model species and will allow identification of markers that can be assessed using targeted techniques for accurate age estimation in large population studies.

4.2 Introduction

Understanding the age structure of populations is a key aspect of animal ecology and conservation. Age estimate information can help to determine animal mortality, susceptibility to parasites, reproductive life history and the impact of anthropogenic activities (Froy *et al.* 2013; Gianuca *et al.* 2017; Musick 1999; Scott 1988). However, measuring the chronological age of many wild animals is a difficult task due to the lack of external changes that reflect age. Some animals have quantifiable physical changes as they increase in age, for example, tooth length in deer (Pérez-Barbería *et al.* 2014) and growth rings in fish otoliths (Buckmeier *et al.* 2002; Campana 2001; Gunn *et al.* 2008). However, few of these can be measured without capturing or even killing the animal. The impact and ethics of these interventions on animals is often the subject of debate (Festa-Bianchet *et al.* 2002; Nelson 2002). Other animals can show general changes with life stage, for example, plumage variation in some seabirds (Weimerskirch *et al.* 1989); or larval stage of arthropods and molluscs (Cobb & Wahle 1994; Ernande *et al.* 2003), but these often only provide age information for immature individuals. This lack of accessible chronological age information limits our understanding of many wild animal species and it is only through long term, expensive tracking or marking studies that age data can be collected and used effectively.

Molecular biomarkers of age have recently been the focus of an increasing number of studies (Ito *et al.* 2018; Maegawa *et al.* 2017; Wright *et al.* 2018). Neither telomere length or DNA damage markers have been successfully used for chronological age estimation in a wild animal population, so there is interest in developing alternative molecular age biomarkers (Dunshea *et al.* 2011; Jarman *et al.* 2015). One promising avenue is measuring epigenetic modification controlling changes in gene expression that occur during animal ageing. Epigenetic regulation of gene expression can occur at several different levels and can include histone modification, non-coding RNA (ncRNA) and DNA methylation (DNAm). DNAm, the addition of a methyl group to a cytosine followed by a guanine (CpG site), has been examined in the most detail and recent evidence supports the use of this epigenetic modification for individual age determination (Hannum *et al.* 2013; Horvath 2013; Vidal-Bralo *et al.* 2016).

Here, we refer to two types of changes in DNAm with age that could be used to estimate age in wild animals. 'Epigenetic drift' generally refers to broad DNAm signals at sites distributed across the genome, which in mammals, birds and fish has been reported to decline with age (Gryzinska *et al.* 2013; Jakubczak *et al.* 2016; Shimoda *et al.* 2014). Drift signals can also be enriched in CpG islands and enhancers (Slieker *et al.* 2016). 'Clock-type' markers are specific CpG sites that show a strong correlation with known chronological age. Correlations observed in this category can be tissue specific and can involve an increase (hypermethylation), or decrease (hypomethylation) with age (Horvath 2013; Slieker *et al.* 2018). Clock-type CpG age markers have recently been referred to as "age-related DNA methylation positions" (aDMPs) (Lowe *et al.* 2018; Slieker *et al.* 2018). aDMPs are generally located within the promoter or first exon of a gene (Bekaert *et al.* 2015; Grönniger *et al.* 2010; Horvath 2013; Sziráki *et al.* 2018; Zbieć-Piekarska *et al.* 2015a). Epigenetic drift is thought to occur due to a decline or imperfect replication of DNAm by an epigenetic maintenance system with increasing age (Horvath 2013; Horvath & Raj 2018). However, the mechanisms for specific 'clock-type' aDMP change have not yet been characterised.

Very little is known about DNAm in most non-model species, especially birds. Available studies have mostly focused on model species such as the Red junglefowl (*Gallus gallus*) (Gryzinska *et al.* 2013; Hu *et al.* 2013b; Li *et al.* 2011) and Japanese quail (*Coturnix japonica*) (Andraszek *et al.* 2014). These studies show a distribution of DNAm in the genome similar to that observed in mammals. Epigenetic drift is the only age-related DNAm change that has been reported in birds. Gryzinska *at al.* (2013) observed DNAm changes between chickens aged between 1 day and 32 weeks using a colorimetric immunoenzymatic based protocol. We have previously reported that the DNAm status of several mammalian clock-type age-related genes were not conserved in homologous regions of a seabird (De Paoli-Iseppi *et al.* 2017b).

Here, we used known age individuals from a long-term study of Short-tailed shearwater (*Ardenna tenuirostris*) to investigate age related changes. The shearwater has high breeding site and partner fidelity and is long-lived, making it an ideal species in which to study population status and chronological ageing in a seabird population. Fisher Island (Tasmania, Australia), is the site of a long-term banding study of this species and as such can be used to

collect known age blood and feather samples for the investigation of DNAm and chronological age (Bradley *et al.* 1991). Epigenetic age estimates of seabirds would be particularly valuable for use in population viability analyses and could further our understanding of environmental effects on animal performance or foraging (Velarde & Ezcurra 2018). For the first time, we have used digital restriction enzyme analysis of methylation (DREAM) to assess DNAm in a non-model vertebrate. We identified seven aDMPs in DNA extracted from 71 whole blood samples. A model relating methylation at these aDMPs to age was made and the precision evaluated using the mean absolute difference (MAD) between the estimated and known chronological ages. Our study is the first to identify DNAm changes with chronological age in a wild seabird and will provide a foundation for further study of age-related DNAm in non-mammalian vertebrates.

4.3 Methods

4.3.1 Samples and DNA extraction

In sampling trips between 2015 – 2018, blood samples were collected from adult (November – December) and chick (March) *A. tenuirostris* from Fisher Island (40°13'00.7"S 148°14'20.7"E) Tasmania, under Department of Primary Industries, Parks, Water and Environment (DPIPWE) permit: FA15230 and University of Tasmania (UTAS) Animal Ethics Committee permits: A14277 and A0016107. Blood was collected onto Whatman FTA® Micro (WB120210) cards and stored as previously described (De Paoli-Iseppi *et al.* 2017b). DNA was extracted from a 3 mm punch of immobilised blood using an Epicentre MasterPure™ (MCD85201) DNA Purification Kit according to the manufacturer's instructions. We examined blood DNA in two high-throughput sequencing runs of a total of N = 71 known-age individuals. Age was determined by recording the band number of birds first marked as chicks, and was rounded to whole years as all sampling occurred in a short time window each year. Run 1 consisted of 35 known-age animals (5 – 21 years old, mean = 12.14 years). Two individuals aged 8 and 14 years old were replicated within this run. Run 2 consisted of DNA from 36 additional known-age samples (6 – 21 years old, mean = 14.18 years). Run 2 contained three technical replicates from Run 1 (6, 12 and 21 years old) and three within-run replicates aged 8, 14 and 21 years old. Several birds were recaptured in sampling trips in different years allowing us to perform some limited longitudinal observations (Run 2: N = 3 x 2 samples and N = 4 x 2 samples at 1 and 2 year resights

respectively). In total, *N* = 63 known-age shearwater were used to calibrate the model following removal of replicates. Bird sex was determined by *CHD-1* gene amplification in blood DNA using a previously described method (Faux *et al.* 2014). Sample details for each age group and known age distribution are shown in Table 1 and Supplementary Figure 1 respectively.

Table 1. Short-tailed shearwater (*A. tenuirostris*) sample details.

Group	Age range (years)	Samples	Longitudinal data	Male	Female	Replicates
Chicks	0.15	2	-	1	1	-
Young breeders	5 - 9	19	-	8	9	WR: *N* = 2, BR: *N* = 1
Middle age	10 - 18	42	6	24	18	WR: *N* = 2, BR: *N* = 1
Old	19+	9	1	3	6	WR: *N* = 1, BR: *N* = 1
Total	0.15, 5 - 21	71	7 pairs^	36*	34	WR: *N* = 5, BR: *N* = 3

* 1 sample failed sexing assay (depleted DNA)

^ Refers to 14 samples

WR: within run, BR: between run

4.3.2 Analysis of genome-wide 'CCCGGG' methylation

We examined DNAm at CpG sites throughout the genome using digital restriction enzyme analysis of methylation (DREAM) of 71 Short-tailed shearwater whole blood DNA samples (Jelinek & Madzo 2016). Briefly, genomic DNA (1 μg) extracted from shearwater blood FTA samples was sequentially cut with two enzymes that recognise the 'CCCGGG' sequence motif in DNA (Figure 1). Methyl-sensitive *SmaI* first cuts only unmethylated sites leaving blunt 5'-GGG ends. Then, *XmaI* cleaves the remaining methylated sites leaving 5'-CCGGG ends. Thus, unique sequences are made for methylated or unmethylated CpG sites. Following this sequential digest, DNA was used to create sequencing libraries using NEBNext Multiplex Oligos for Illumina Index Primer Sets 1 – 3 and standard Illumina protocols. Blunt-end ligation is done using NEBNext adaptor (10 μm) and T4 DNA ligase with hairpin loop cleavage with USER enzyme. Dual size selection for 250 – 450 bp fragments was done using AMPure XP beads. Unique barcodes were then added to DNA from individual samples with 12x rounds of PCR using AmpliTaq Gold DNA Polymerase (see Supplementary Table 1). Individual barcoded samples were analysed for correct library size distribution (250 – 450 bp) using high sensitivity DNA 1000 kits on the Bioanalyzer 2100. Two microliters of each sample was also quantified using a Qubit 2.0 to ensure equal volumes were pooled in the final library. Libraries were run at 2 – 4 ng/uL on the Illumina NextSeq 500 platform with a 15 – 25% PhiX control at the Ramaciotti Centre for Genomics (UNSW, Sydney, AUS).

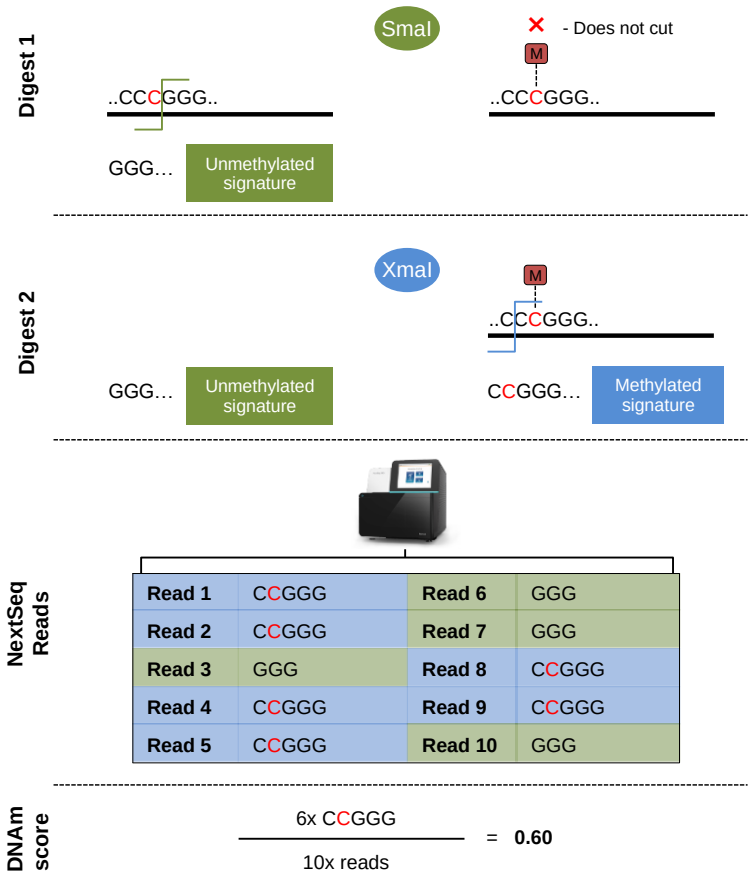


Figure 1. Digital restriction enzyme analysis of methylation (DREAM). A schematic showing an example of sequential digestion and sequencing of a single 'CC**CG**GG' site for one sample. In digest 1, a methylation sensitive enzyme (*SmaI*), is used to generate unmethylated signatures (GGG). *SmaI* does not cut methylated cytosines (methylated cytosines are indicated by red text and a floating, red 'M' box). Following this, *XmaI* is added to the sample in digest 2, and generates methylated signatures (C**CG**GG). All samples are then run on a next-generation sequencing platform (e.g. Illumina NextSeq), and read counts of each signature are counted. In this example, for ten reads of a unique CpG marker, six contain the methylation signature giving a DNA methylation (DNAm) score of 0.6.

4.3.3 Statistical analysis and construction of an age prediction model

4.3.3.1 Sequencing data analysis pipeline

Raw DNA sequence reads were run through an in-house data analysis pipeline in the following steps.

1. *Quality filtering*. Demultiplexed Fastq sequences were filtered with a maximum expected error (maxee) rate of 0.5 and converted to Fasta format (Edgar & Flyvbjerg 2015).
2. *Dereplication*. A database of unique reads from all samples was generated (dereplication) using trimmed sequences and the USEARCH10 command 'fastx_uniques' (Edgar 2010), with a min_unique size = 150.
3. *Methylated and non-methylated motif databases*. These dereplicated sequences were duplicated to contain the unique sequence with either the 5'-GGG or 5'-CCGGG motif, in separate databases (GG or CC databases).
4. *Motif database hits*. Each sample was then compared to each database using the 'usearch-global' command with 97% identity and required an exact match to the first 2 bp of the relevant motif (id_prefix = 2). Hits for each sequence against both methylation databases were recorded.
5. *DNAm level calculation*. The methylation level for each sample was then calculated as the count of the methylated signature divided by the total number of hits for a specific CpG marker and the value was recorded between 0 and 1. A value of 0 is unmethylated (i.e. all sequences from that site match the GG sequence generated by methyl-sensitive *SmaI*) and 1 is methylated (i.e. all sequences from that site match the CC sequence generated by *XmaI*).

Methylation scores were retained for read depths between 20 and 2000 reads. Scores that were calculated outside of this range were converted to a 'NA'. To retain potentially informative markers in the final analysis, markers with less than seven NA values across all samples were imputed using the mean of the remaining non-NA values for the marker. This method ensured that potential age-related markers would not be omitted based on missing scores and that imputed values would have a relatively small effect on any correlations observed. Since variation is required to find correlations with age, we removed markers that had a DNAm standard deviation of less than 5% across all samples. A small run effect was

observed, so the mean DNAm difference between run 1 and 2 replicates was used to adjust the score of each marker in run 2.

4.3.3.2 Predictor selection and age estimation model

Markers that passed filtering were then used to fit penalised lasso regularisation paths to each predictor using the R package 'glmnet' (Friedman *et al.* 2010). The penalty value used to select coefficients, lambda 1 standard error (λ1se), was calculated after repeated runs (100x) of the default k-fold cross validation function of glmnet (cv.glmnet, 10-fold) with an alpha = 1 (lasso). This method randomly subsets the data each cycle and assesses the linear relationship between age and DNAm. Following repeated runs of this function a mean λ1se value was generated. The λ1se value generally selects CpG sites for the simplest model with an error similar to the best model (λ minimum), given the cross-validation uncertainty.

Individual markers that passed the λ1se cut-off were inspected visually using simple linear regression and markers that had an $R^2 < 0.2$ or showed small changes in DNAm range (< 15%) were removed from further analysis. Remaining age-related CpG sites were then incorporated into a multiple linear regression model. To test the selected markers, the original data set was randomly split into 75/25% training ($N = 47$) and test ($N = 16$) data sets respectively. Training set DNAm values for each aDMP were used to create a multiple linear regression model. The model was then tested with remaining samples in the test set. This random sub-sampling method was run for 100 iterations. By substituting the calculated methylation values for each of the individual shearwaters used in the training and test sets into the equation, we obtained the predicted epigenetic age. Mean absolute difference (MAD), the uncertainty of age estimates expressed in years, between the known and estimated age was then calculated. The 77 bp sequence following the CG motif was analysed by BLASTn searches of bird genomes available on the NCBI database to identify any regions conserved between species (Altschul *et al.* 1990).

4.3.4 Global DNA methylation analyses

4.3.4.1 Global analysis of 2338 CpG sites using DREAM

The mean DNAm of 2338 CpG sites identified using DREAM were analysed by age group in years as follows: Chicks: 0.12 – 0.15 (N = 2), Young breeder: 5 – 9 (N = 16), Middle: 10 – 18 (N = 39), Old: 19+ (N = 6). CpGs were analysed using a one-way ANOVA followed by post-test for multiple comparisons (Tukey's HSD). Mean DNAm differences were calculated in both the chick and young breeder context and analysed as above. Significance was set at $P < 0.05$.

4.3.4.2 Colorimetric DNA methylation analysis

We also measured epigenetic drift in global DNAm using a commercially available methylated DNA quantification assay for relative 5-mC content (Abcam, Colorimetric, ab117128). Briefly, 42 shearwater blood DNA samples (chicks, 5 – 21 years old, mean = 10.9 years) were analysed in duplicate, alongside the supplied positive (5 ng) and negative controls. Methylated DNA was captured and detected using diluted (1:1000, 1:2000) 5-mC antibodies. Following the addition of a developing solution, colour change was monitored and quantified at 450 nm (Tecan Spark). Using the mean absorbance values of the duplicates, relative 5-mC for each sample was calculated as follows: (((Sample OD – Negative control OD) / DNA input (ng)) / (((Positive control OD – Negative control OD) x 2) / Positive control input (5 ng))) * 100. Analysis of duplicate colorimetric data was done using a one-way ANOVA with Šidák correction for multiple comparisons for each age group in years as above.

4.4 Results

4.4.1 Sequencing metrics

Quality analysis of DREAM libraries showed bands in the expected post clean-up range, (range = 194 – 974 bp, mean = 451 bp; Bioanalyzer gel and electropherogram traces are shown in Supplementary Figure 2A – D). A total of 125 million sequences (mean of 1761622 per sample) passed initial bioinformatic QC (maxee = 0.5 and matched restriction site motif; Supplementary Table 2). The sum of reads from sequences with a mean high read depth (> 2000x) represented approximately 6% (mean = 84518 reads) of the total mean sequences per sample. Following filtering and dereplication, we identified 93884 unique sequences that

were used to create a database of reference sequences (i.e. markers for specific CpG sites) for sample matching (Supplementary Figure 3). Following the pipeline filtering described, a total of 2338 unique CpGs were used for lasso analysis (glmnet).

4.4.2 Development and testing of an age prediction model in the Short-tailed shearwater

DNAm data from seven CpG sites obtained using DREAM were included in the age prediction model based on our selection criteria (Figure 2A – G). Information on removed CpG sites with weaker age correlations is provided in Supplementary Table 3 (e.g. just below our mean λ1se cut-off of 1.2; see Supplementary Figure 4). To investigate potential sex-related DNAm effects in the seven aDMPs used in the age prediction model, separate linear regressions were done for each sex (Supplementary Figure 5A-G). Sex had a significant effect on DNAm age correlation in a single aDMP in isolation (M1801, $P = 0.0031$, Bonferroni corrected), with males driving the association (Supplementary Figure 5C). However, there was no sex-specific effect when the methylation scores for all seven aDMPs were then used to create the age estimation model (Figure 3, sex regression slopes and diagnostics are shown in Supplementary Figures 5H and 6 respectively). Read depth had a mean of 51x for these CpG sites (Supplementary Figure 7). The MAD between the known and estimated age reports the uncertainty in age estimates expressed in years. Following repeated cross-validation, the seven aDMP age assay provided epigenetic age estimates in training subsamples with a MAD of 2.34 ± 1.73 (SD) years (mean $R^2 = .605$, range: 0.46 – 0.72) (Supplementary Figure 8A). In the validation test subsamples, the age estimates had an increased error; across all age estimates MAD = 2.81 ± 2.08 years (mean $R^2 = .404$, range: 0.03 – 0.80) (Supplementary Figure 8B). The significant y-intercept of 5.13 indicated that the predicted ages were overestimated for chicks and young birds and underestimated for older individuals, and may indicate a non-linear relationship. The training set MAD ranged from 1.17 – 6.25 years, whilst in the test set MAD ranged from 1.58 – 7.86 years. The MADs for each year and grouped age, as described in the methods, are shown in Figure 4. Between run replicates for seven age-related CpG sites showed a mean DNAm score difference of 11.29% (range: 3.79 – 12.80%) and 6.83% (range: 4.21 – 11.04%) pre- and post-run adjustment respectively (Supplementary Table 4). Within run replicates showed a mean absolute difference in DNAm of 8.65% (range: 5.18 – 11.91%) for the age-related markers.

4.4.3 Biomarker sequence and gene conservation

The seven aDMPs we identified were used to search for conserved regions in available bird genomes and scaffolds using BLASTn. Of these seven markers, four had low E values and > 50% query cover indicating a reasonable match with a known sequence in the available avian databases (Table 2). Marker 1071 matched with the *G3BP1* region in the Zebra finch (*Taeniopygia guttata*) genome, however the query cover was only slightly above 50%. Marker 1934 had a 100% query cover match with an uncharacterised locus in the Mallard (*Anas platyrhynchos*) genome. Marker 2083 matched against scaffold 4695 in the North Island brown kiwi (*Apteryx australis mantelli*) genome. Finally, marker 3169 had a 100% query cover match to the *DHH* gene in several species with the top hit to the Eurasian blue tit (*Cyanistes caeruleus*) genome.

4.4.4 Longitudinal observations of DNA methylation in resighted individuals

We observed that 6/7 (85%) age estimates for resighted individuals sampled 1 or 2 years apart showed the expected positive increase in predicted age relative to their known age from leg bands (Figure 5). At many individual aDMPs the longitudinal samples did not follow the expected DNAm trend (Supplementary Figure 9A-B). However, when combined into the model, only one individual showed a negative change in estimated age from two samples taken at 15 and 17 years of age. The mean absolute difference between estimated and known age for 2-year resights was 0.74 years (N = 8) and 0.87 years (N = 6) for 1-year resights.

Table 2. Age-related CpG site sequences and BLASTn results (birds only).

CpG	Sequence (Illumina NextSeq 77 bp)*	Adj. R^2	DNAm range (%)	DNAm direction	Sequence GC content (%)	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
1071	**CCCGGG**GAACATAACCAGGGCCCGAGGAACTGAACCAGAGCTCCAGGAACAAAACCAGAGCCAGAGGAACTAAACAA	0.258	0.32	Pos	54.5	*Taeniopygia guttata (Zebra finch)*	G3BP1 (mRNA)	51	86	4.10E-02
1158	**CCCGGG**GAAGCTGCCAAAAGCGCGAAGCCGCTGGCGGACGCGGAGGTAAGAGCTGAGAGGGACGTGCCGGTGCCTCG	0.290	0.74	Neg	70.5	*Apteryx australis mantelli (North Island brown kiwi)*	Genomic scaffold 187	46	86	1.40E-01
1801	**CCCGGG**AGATGCCGGGAAATGTAGCCGTGCCGCGGCTGGCCCATCCGGCAGCGCCGTCCCCCGCTCTGCCGGCGTCT	0.285	0.55	Pos	75.3	*Lonchura striata domestica (Society finch)*	AKAP10 (mRNA)	42	88	1.40E-01
1934	**CCCGGG**TTTGAATTACTATTGAATAAGCAGCAATGAAATCTCTATCAAAATAATCAGTACTTCCAAAAACCACAAAC	0.236	0.35	Neg	33.8	*Anas platyrhynchos (Mallard)*	Uncharacterised LOC106014977 (ncRNA)	100	97	1.00E-28
2083	**CCCGGG**CCCAGGGCCAGCTGCCGGGCTTGGCCCCAGGAGGAGGAGGAGAAGGAGGAGGAAGGCAGGATCTCCAAGGC	0.258	0.89	Neg	71.4	*Apteryx australis mantelli (North Island brown kiwi)*	Genomic scaffold 4695	64	84	8.00E-05
3169	**CCCGGG**CCGGGGCCCGGTGGGGCGGCGGCGGTTGGGGCGGCGGCAGCTCTCGCCGCTGCTCTACAAGCAGTTCGTGC	0.251	0.55	Pos	79.2	*Lepidothrix coronata (Blue-crowned manakin)*	DHH (mRNA)	100	99	8.00E-30
3784	**CCCGGG**GACCGGGTGCACATCTGGGCTGAGAGATGCTGAAAGCAAAGAGCAGCCGGAGCGCTGGGAGAGCGGTGGGA	0.211	0.36	Pos	66.2	*Cyanistes caeruleus (Eurasian blue tit)*	ANAPC16 (mRNA)	35	93	1.40E-01

*Restriction site is shown in bold with CpG sites underlined.

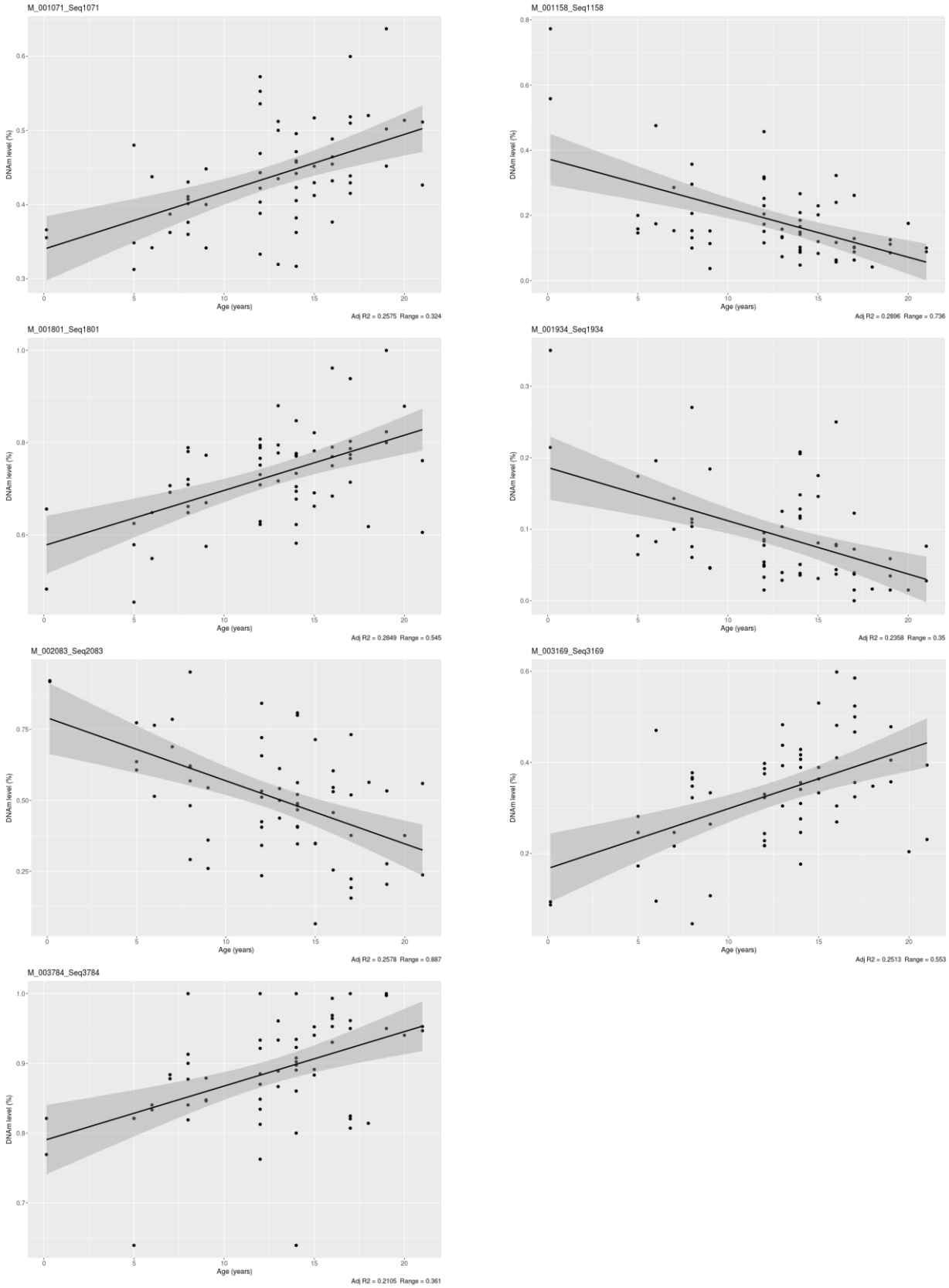


Figure 2A – G. DNA methylation of selected CpG sites showing a relationship with chronological age. Linear regression of DNA methylation and chronological age for each CpG selected using lasso penalisation from a total of *N* = 63 Short-tailed shearwater blood samples. Sequence details for each CpG are shown in Table 2.

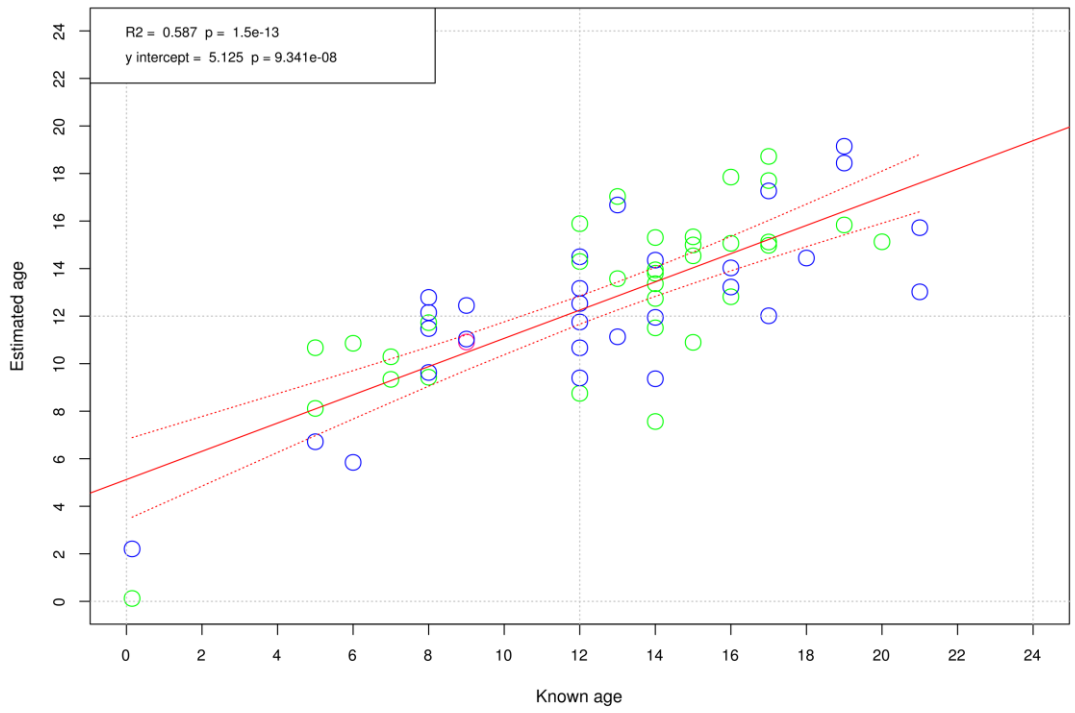


Figure 3. Full multiple linear regression model. Multiple linear regressions for predicted ages of all *N* = 63 Short-tailed shearwater from quantification of CpG methylation at seven CpG sites. 95% confidence limits of the placement of the regression line are shown. Females are shown in blue (*N* = 30), males in green (*N* = 32), and a single unknown sex (*N* = 1, 9 years old) is shown in pink.

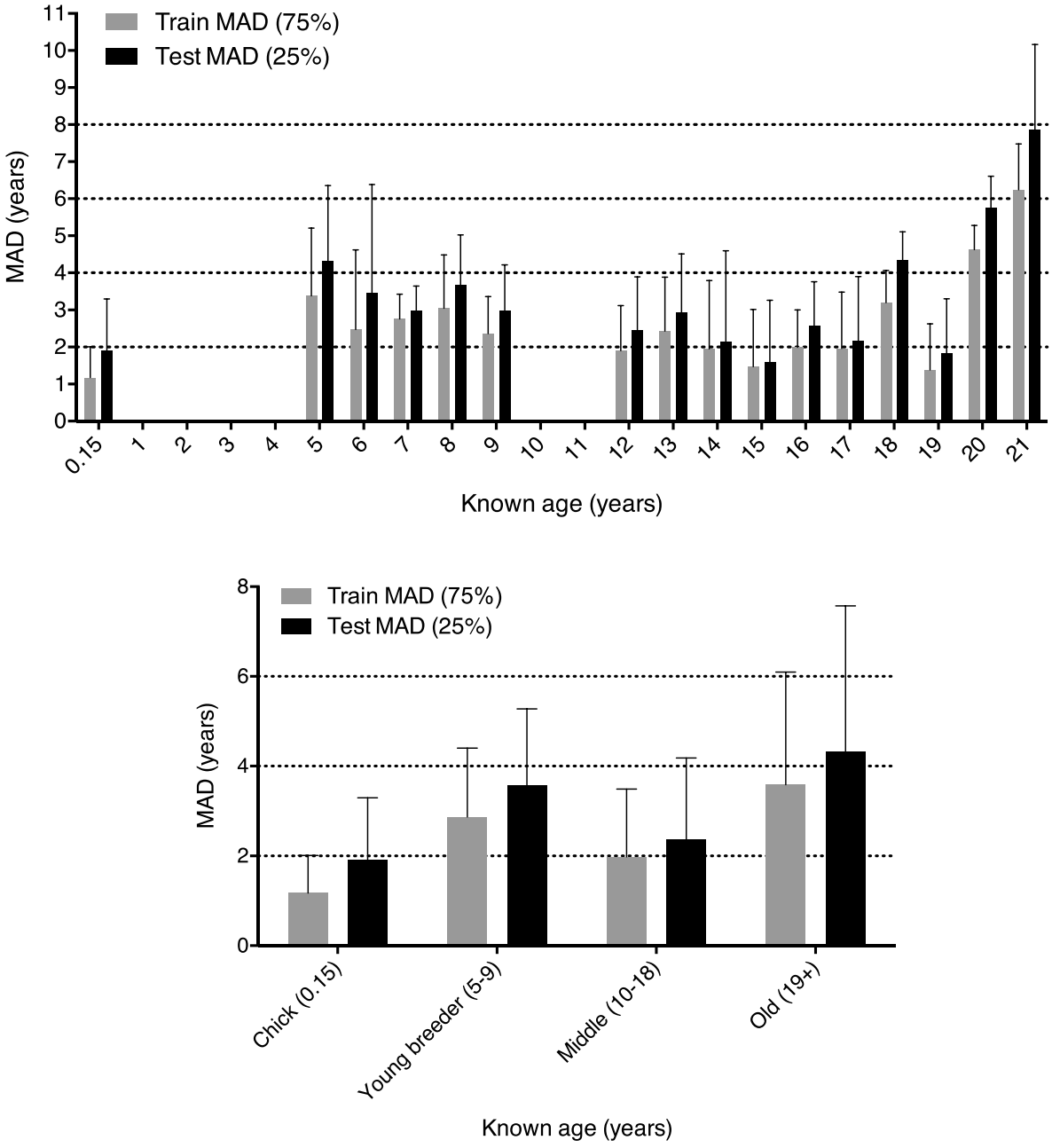


Figure 4. Yearly and age-class grouped mean absolute deviation (MAD). Determined by the absolute difference between the estimated and known age, the MADs are shown **(A)** for each year of age for known-age animals included in the model and **(B)** for each of the described age groups.

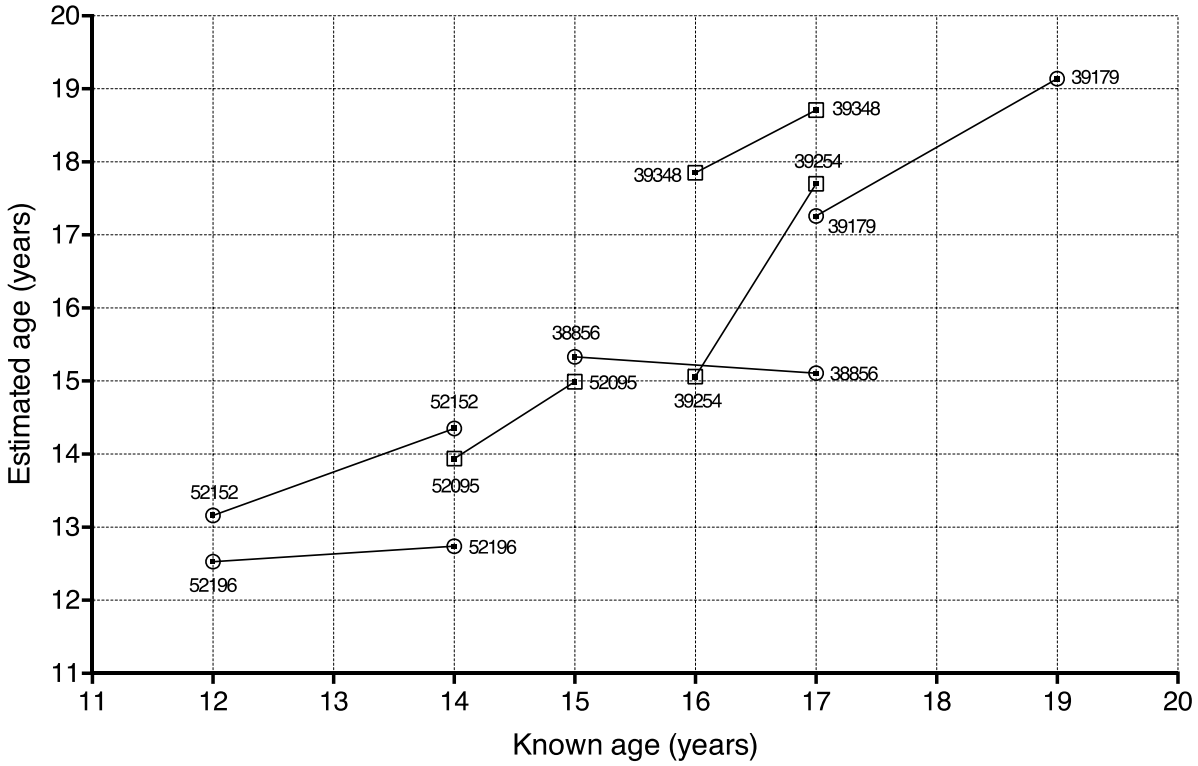


Figure 5. Longitudinal DNAm data for 1 and 2 year resights. The estimated epigenetic age versus known age (from leg bands) for each individual bird with longitudinal resights (1-year *N* = 3; 2-year *N* = 4).

4.4.5 DNA methylation of 2338 CpGs using DREAM assay

We show that a large proportion of the 2338 CpG sites that passed the filtering cut off are highly methylated, with 50.2% of CpGs showing DNAm levels greater than 80% across all ages (Figure 6A). We also observed a small, but non-significant linear change in DNAm from young animals to old. The mean DNAm was .712, .724, .725 and .729, for chicks, young breeders, middle and old birds respectively (Figure 6B). The difference in mean DNAm, relative to chick levels, for each individual CpG site is shown in Figure 6C. This shows that relative to older birds, chicks are less methylated at low DNAm levels (approximately < 10%) and more methylated at high DNAm levels (approximately > 90%).

4.4.6 Global 5-mC using colorimetric assay

Relative 5-mC was quantified against the supplied 5 ng positive control. Global blood DNAm levels of the Short-tailed shearwater were combined into age groups as described in the methods. Chicks and young breeders showed similar relative 5-mC levels, (mean = 0.725, N = 4 and mean = 0.727, N = 15 respectively). Both of these groups had slightly higher relative 5-mC than that observed in middle-aged birds (mean = 0.614, N = 17) and old birds (mean = 0.498, N = 5). Following adjustment for multiple comparisons, no significant differences were observed between the age groups (Figure 7).

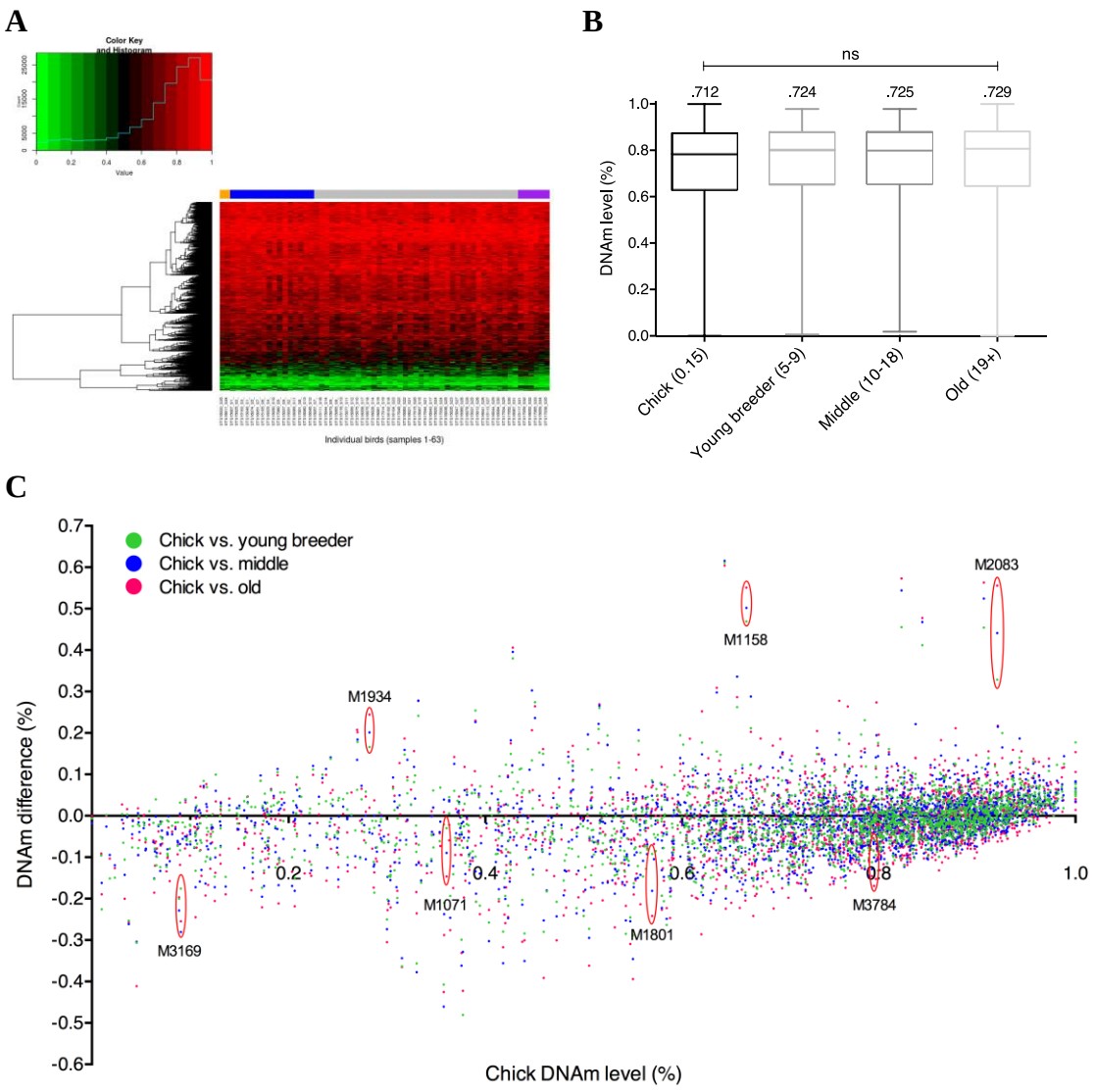


Figure 6. DNA methylation in chicks, young, middle and old shearwater. A. Heatmap of 2338 CpGs analysed using DREAM in shearwater. Age groups are defined by column header colours, chicks are shown in orange, young breeders (5 – 9 years) in blue, middle (10 – 18 years) in grey and old (19+ years) in purple. A colour key indicates the DNAm value ranging from 0 (unmethylated, green) to 1 (methylated, red). **B.** The mean DNA methylation (DNAm) level of each of 2338 CpG sites (N = 63 birds) for each age group: chicks (N = 2), young breeders (5 – 9 years, N = 16), middle aged (10 – 18 years, N = 39) and old (19+ years, N = 6). A post-ANOVA test for linear trend did not indicate that the small increase in DNAm was significant. **C.** The difference in mean DNAm for each CpG in a chick context (high to low) versus older age groups, i.e. a positive value indicates higher DNAm in chicks and vice-versa. Age-related CpGs used in the model are circled in red and are labelled as shown in Figure 2.

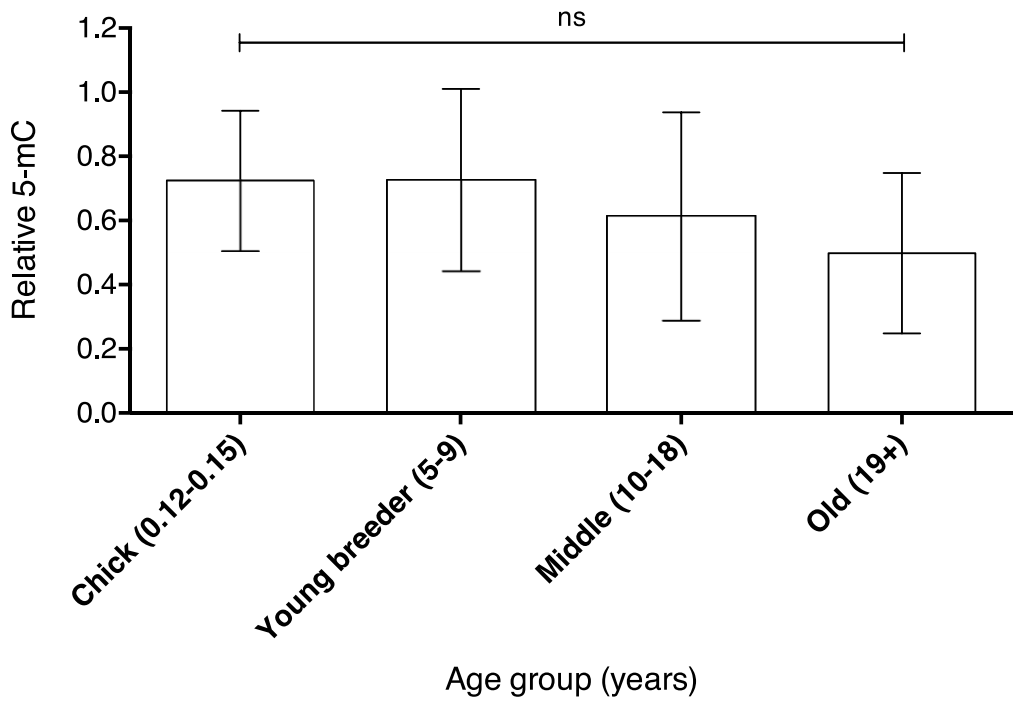


Figure 7. Global 5-mC of known-age shearwater using colorimetric assay. Relative global 5-mC content was assayed using a colorimetric based assay and analysed with a plate reader. A trend towards global hypomethylation was observed in $N = 42$ known-age Short-tailed shearwater, however this result was not significant following correction for multiple comparisons ($P > 0.05$).

4.5 Discussion

Seabirds exhibit little or no external physical changes with age and there are currently no reliable biomarkers of chronological age in most long-lived seabirds beyond fledging. The identification of an accurate age biomarker would be a substantial advance in our ability to understand seabird age-related demographics. Seabird age estimation using molecular methods is currently not possible. DNAm changes with age have been reported for both wild and model mammalian species in several tissues, indicating that DNAm age biomarkers may be useful in birds. In this study, we quantified the DNAm profile of known-age Short-tailed shearwaters using digital restriction enzyme analysis of methylation (DREAM). We present evidence for DNAm changes with chronological age in seven CpG sites.

4.5.1 Age related biomarkers in birds

Previous bird ageing research has focused primarily on telomere length assays and pentosidine accumulation in collagen. Studies of terminal telomere restriction fragments (TRFs) have shown that telomere length can shorten with increasing age and that the rate of change corresponds to lifespan in several species (Bize *et al.* 2009; Juola *et al.* 2006; Tricola *et al.* 2018). However, this trend is not consistent amongst all birds, with some species showing increases in TRF with age, as in the Leach's storm-petrel (*Oceanodroma leucorhoa*) (Haussmann *et al.* 2003), and no decline in length, or both as reported for the Magellanic penguin (*Spheniscus magellanicus*) (Cerchiara *et al.* 2017). For individuals in some avian species, change in telomere length can be tracked longitudinally and correlate with reproductive timing, however the use of TRF for cross-sectional analysis of age has yet to be demonstrated (Bauer *et al.* 2018).

Pentosidine is a less frequently studied age biomarker for birds. It forms cross-links between amino acid residues in collagen and accumulates with age in birds (Fallon *et al.* 2006; Iqbal *et al.* 1999). Pentosidine has been shown to accumulate in a linear fashion in terrestrial birds and some seabirds including California gulls (*Larus californicus*) (Chaney Jr *et al.* 2003) and Double-crested cormorants (*Phalocrocorax auritus*) (Fallon *et al.* 2006). This technique has yielded age estimates with a precision of 2 – 4 years in wild birds (Chaney Jr *et al.* 2003; Fallon *et al.* 2006; Rattiste *et al.* 2015). However, in a study of another long-lived seabird, the Bridled tern (*Onychoprion anaethetus*), no correlation between pentosidine levels and age

was found (Labbé 2017). It is not known how pentosidine levels may respond to the effects of changing biological age or environmental stressors. As a result of the limited success in age estimation by these methods, our research aimed to build upon recent successes in mammals by assessing DNAm estimates of age in the Short-tailed shearwater.

We previously established that specific aDMPs from mammals were not conserved in the shearwater (De Paoli-Iseppi *et al.* 2017b). We therefore sought to identify bird-specific aDMPs or a global DNAm signature associated with age using DREAM of whole blood samples. This is the first epigenetic age assay developed for use in a seabird, and one of the few used in a wild species. Using the DREAM method, we identified seven novel aDMPs in shearwaters. Following repeated cross-validation of our known-age samples to train and test the age-estimation model, we reported a test-set MAD for all ages of 2.81 ± 2.08 years. The linear relationship with age in these CpG sites is not as strong as those reported for whales (Polanowski *et al.* 2014) or dogs (Thompson *et al.* 2017), but was similar to that reported for a bat species (Wright *et al.* 2018). We also observed variation in MAD for different age classes, with birds aged 5 – 9 years and 19+ providing less accurate age estimates compared to other groups (Figure 4A). Additionally, the significant Y-intercept in our model (Figure 3) causes an overestimation of age in younger individuals. A single marker (M1801) showed evidence for male driven DNAm age correlation. Due to the reduced sample size when comparing by sex only, more known-age samples would be required to confirm the lack of association in females and ideally, whole genome information could determine if this marker is located on a sex chromosome.

However, the biggest limitation in developing our model was the low number of young non-breeding bird samples that we could capture in the field, which hinders our understanding of the rate of DNAm change between chicks and early breeders (5 – 9 years), and with more samples, this may be correctable in future. The shearwaters studied here typically do not return to their island of birth until their first year of breeding at age five (Bradley *et al.* 1991; Bradley *et al.* 1989). However, for unknown reasons we did not recover many individuals in the 5 – 9 early breeder age range. The larger DNAm variability in these young animals could be due to the stressful effects of the first year of breeding. Shearwaters lay one of the largest eggs relative to body mass of all seabirds, and individuals face challenges including

incubatory fasting and intermittent foraging (Wooller *et al.* 1990). Additionally, both migration and parenthood can reduce body condition, and evidence suggests that these birds may undergo intermittent breeding if an individual determines its body condition is too low (Bradley *et al.* 2000b).

Despite some uncertainty in ages estimated with our model, this approach could discriminate between relevant age classes (e.g. young and old adults). These epigenetic age estimates, in combination with other parameters including sex and weight, could be used to examine the effect of climate change on population viability (Lee 2017). Recent studies also highlight other areas in which estimated age data could be informative, including post-pest eradication monitoring of island-breeding seabird populations (Brooke *et al.* 2018), parasite load in the Blue tit (Aguilar *et al.* 2016) and modelling the impacts of longline fisheries on effective population size (Cortés *et al.* 2018; Mills & Ryan 2005).

Obtaining a broad age range of samples from long-lived, known-age birds is difficult as extensive banding studies are rare. Whilst the Fisher Island shearwater population has been followed for several decades, the youngest and oldest adult individuals we recovered were 5 and 21 years old respectively. The oldest individual, at 21 years old, represents a little over half of the maximum reported lifespan for this species of 39 years. However, research on age dependent survival on Fisher Island birds shows few animals living beyond 25 years post first breeding, which would place our oldest individual at closer to 70% of the expected lifespan of approximately 30 years (Baylis *et al.* 2018; Bradley *et al.* 1989). The relationship we have observed with age should be investigated further for older individuals, however previous studies in mammals have primarily shown linear correlations with age (Maegawa *et al.* 2010; Polanowski *et al.* 2014; Spiers *et al.* 2016). Although no recaptures were made within the 1 – 4 year age range, as these non-breeding birds are not at the nesting sites, the relationship of adults to the DNAm level of the chicks suggest birds at these ages will have a similar trend to the rest of the calibration range.

We quantified ‘epigenetic drift’ in DNAm levels observed across all 2338 CpG sites included in our analysis. We did not identify a significant trend with chronological age. However, we

did observe some interesting differences between young and old age groups at the lower and upper limits of DNAm. In contrast to mammalian and the only other bird study, we found no clear trend of DNA hypomethylation in older animals compared to that in younger individuals (Gaudet *et al.* 2003; Gryzinska *et al.* 2016; Portela & Esteller 2010). The lack of statistical significance could be due to the analysis of this relatively small subset of total CpGs in the bird genome.

Immunoenzymatic analyses of chicken 5-mC levels have shown decreased global methylation with age (Gryzinska *et al.* 2013). Using the same method, we found no relationship between relative 5-mC levels and age in 42 known-age shearwater whole bloods. However, we observed a non-significant trend towards decreasing methylation across age groups. Our study of age-related global DNAm in shearwaters is only the second of this phenomenon in birds and further work will be required to determine if this approach could be suitable for age estimation in other bird species.

4.5.2 Measuring methylation in non-model organisms

Despite the identification of several thousand unique CpG sites using the DREAM method, the 20x read depth requirement for DNAm calculation resulted in the exclusion of many sites from further analysis. A small percentage of the total reads was also lost to repetitive elements. There is little doubt that as technologies improve sequencing depths will increase, and direct analysis of CpG DNAm will be possible, (Rand *et al.* 2017; Slatko *et al.* 2018). Improvements in bioinformatics will also help to validate DNAm markers and predict age in large data sets (Vidaki *et al.* 2017). The DREAM technique has been used previously to identify DNAm changes following compound exposure in zebrafish embryos (Bouwmeester *et al.* 2016) and caloric restriction in mice (Maegawa *et al.* 2017). A similar method, EpiRADSeq, also uses a methylation sensitive restriction enzyme (*HpaII*) and NGS to quantify DNAm in CpG sites (Schield *et al.* 2016). This technique differs from DREAM in that only a single methylation sensitive enzyme is used in combination with a frequent cutter (*PstI*). *HpaII* recognises a ‘CCGG’ motif, which is likely to lead to higher genomic coverage of CpG sites due to increased cut frequency. However, DNAm scores generated using this method are relative to the count of unmethylated EpiRADSeq reads only. This is avoided when using a dual methylation sensitive digest as in DREAM, as reads are generated for both methylated and unmethylated CpGs (Jelinek & Madzo 2016). Reduced representation

bisulphite sequencing (RRBS) can also be used to quantify CpG DNAm, but does require a higher quantity of initial genomic DNA (Meissner *et al.* 2005). The output of these various techniques depends upon several molecular, platform and bioinformatic factors and choices, which is discussed in detail elsewhere (O'Leary *et al.* 2018). Our results now show that the DREAM method can also be used to quantify global DNAm and screen for aDMPs in non-model animals. The primary limitation in applying this method is the high read depth required per CpG site, particularly in organisms with relatively high quantities of repetitive DNA. This makes it cost-prohibitive as a method for applying to population-wide samples, but certainly effective as a screening method for identifying aDMPs. Once aDMPs are identified by DREAM, targeted DNAm scoring assays could be developed to reduce costs for high-throughput applications.

An additional limitation to the simple analysis of shearwater DREAM and indeed most non-model NGS data, is the limited genomic resources available for further analyses. Multiplex restriction site PCR (mRS-PCR) could be used to obtain both up and downstream sequence around an aDMP of interest (Sarkar *et al.* 1993; Weber *et al.* 1998). This method can generate larger reference sequences for use in targeted bisulphite assays such as EpiTYPER, pyrosequencing or other NGS based techniques (Ehrich *et al.* 2005). More sequence information may also result in more accurate comparative genomic analyses against bird genomes that are currently undergoing scaffold alignment. The genes *DHH* and *G3BP1* were identified as conserved age-related sequences from our data and these could be used in future as part of a targeted gene assay in shearwater (Table 2, M1071 and M3169). Whilst we cannot comment on any potential functional effects of DNAm, *DHH* and *G3BP1* encode for signalling molecules in cell morphogenesis and a DNA-unwinding enzyme, respectively. Two other markers also showed high conservation with other bird species, however these hits were either unassigned (M2083) or uncharacterised (M1934). These factors limit our ability to identify biomarkers that have the potential to be used in closely related species, and design a cost-effective, targeted age assay.

4.6 Conclusions

This study demonstrates that seabird age estimates can be generated from a DNA methylation age assay. This minimally invasive method could be used to produce age estimates for Short-

tailed shearwaters from chicks to 21 years old. This is the first time an epigenetic assay has been applied to a wild seabird and could be used in future to estimate population age structure. Further refinement of this method could result in the identification, validation and use of target genes, similar to that in mammals, for related seabird species and see wider use for monitoring and conservation.

4.7 Data and code availability

DREAM count data, adjusted DNAm values for 2338 CpGs, fasta pipeline and variable selection R scripts used in this publication have been deposited in the Dryad Digital Repository at [doi: 10.5061/dryad.n4h3672]. Sample details and raw Illumina sequence data (FASTQ) are available from NCBI/SRA using accession: PRJNA507458, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA507458.

4.8 Acknowledgements

The authors would like to acknowledge the financial support through the Australian Government Research Training Program, the Holsworth Wildlife Research Endowment – ANZ Trustees Foundation and the Joyce W. Vickery Scientific Research Fund – The Linnean Society of New South Wales. We also thank James Marthick, Cassandra Price, Fernando Gonzalez, WILDCARE Friends of Fisher Island, Ross Monash, DPIPWE field volunteers (2015 – 2018) and Tasmania Parks and Wildlife Service Rangers.

4.9 Author contributions

All authors conceived the ideas and designed methodology; RDP, AMP, CRM and MAH collected samples; RDP and AMP did the genetics laboratory work; RDP, BED and SNJ analysed the data; RDP led manuscript writing. All authors contributed to drafts and gave final approval for publication.

Chapter 5 – Bird population age structures estimated by epigenetic analysis enable management of previously unmonitored populations

5.1 Abstract

1. Population age structure is an important aspect of wildlife demography. Age estimates and animal life-tables are commonly used to inform management and to model complex ecological problems involving interactions between target species and humans, other age classes and determining population viability. Banding seabirds as chicks is currently the only way to determine the exact age of most adult seabirds. As most seabird colonies are not actively monitored given the costs associated with annual banding efforts we sought to apply a molecular method of age estimation.

2. DNA methylation (DNAm) has been used to estimate age in several wild mammal species including whales, bats and wolves. We recently showed that age-related differentially methylated positions (aDMPs) in the Short-tailed shearwater (*Ardenna tenuirostris*) can be detected using digital restriction enzyme analysis of methylation (DREAM). Here, we sought to target these aDMPs using a cost-effective and high-throughput next-generation sequencing method to estimate the age of wild un-banded birds.

3. We reveal 14 additional aDMPs in close proximity to the originally selected biomarkers. Using four of these aDMPs we replicated our previous results in a known-age cohort of 109 blood samples (training set: $R^2 = 0.73$). Further testing of the model with a minimum-age cohort ($N = 55$) showed that 39/55 (71%) of these samples were estimated above their known minimum-age between banding and recapture. We also present the first population age structure generated from DNAm estimates for un-banded long-lived seabirds.

4. *Policy implications.* Our study shows that ageing can be tracked through DNAm changes in seabirds. Importantly, our model demonstrated that age estimates were generally above the known minimum-age of birds banded as adults, a common occurrence in colony ringing studies. We also found that epigenetic markers can generate meaningful age structures. Applying the ageing techniques we described could lead to the expansion of monitoring programs with the potential to inform management decisions. This work also provides a foundation for identifying this type of biomarker in any non-model species.

5.2 Introduction

Cross sectional information on the age of individuals is essential to establish and compare different population structures. Chronological age is also used to determine age-specific

mortality rates and fecundity. Knowing the ages of individuals in a population is especially useful for understanding animal life history (Lewison *et al.* 2012). For example, in the wandering albatross (*Diomedea exulans),* a long-lived seabird, experience and reproductive performance generally increases with chronological age through early adulthood, before declining with senescence later in life (Froy *et al.* 2013). However, the latter is not a linear process and there is evidence for a comparatively large increase in reproductive performance at the final breeding attempt (Clay *et al.* 2018; Froy *et al.* 2013). Evidence in *D. exulans* suggests that non-breeding foraging behaviour can impact on the success of the next breeding season (Clay *et al.* 2018). A similar study in Audouin's gull (*Ichthyaetus audouinii*), also found evidence for a food availability effect on variable breeding parameters with age (Oro *et al.* 2014). Mate choice in *D. exulans*, which form long-lasting pair bonds, has also been correlated with age. The authors also identified an influence of human-induced mortality on the sex ratio of individuals within the colony, which may have allowed more mate choice by females (Jouventin *et al.* 1999). Long term ringing data of the Common Quail (*Coturnix coturnix*) has revealed the importance of age structure to the identification and management of high quality breeding areas (Nadal *et al.* 2018). Age structure information has recently been shown to improve predictions on population fluctuation analyses by accounting for age-specific effects of intra- and interspecific competition (Gamelon *et al.* 2019).

Age-related anthropogenic factors have also been described for seabirds, with evidence for age and sex bias in seabird bycatch data (Gianuca *et al.* 2017). These authors call for further data collection on age-class spatial distribution to better understand the impact of fisheries on these species (Gianuca *et al.* 2017). Island dwelling seabirds are particularly susceptible to mortality caused by introduced pest species, and do not always display effective anti-predator behaviour (Sih *et al.* 2010). In philopatric seabirds, benefits are obtained by previous knowledge of the environment, however this can also make them and their offspring reliable prey items (Ekroos *et al.* 2012). A recent study demonstrated increased patch occupation by experienced Audouin's gulls following the introduction of foxes (*Vulpes vulpes*) to known breeding areas. These results support the hypothesis that predator introductions can drive changes in population age structure and drive down reproductive performance through dispersal (Payo-Payo *et al.* 2018). Age estimates could also be useful for determining potential impacts of new energy infrastructure including both onshore and offshore wind turbines and 'wet-renewables' such as tidal turbines and wave energy devices (Furness *et al.* 2012; Green *et al.* 2016).

We previously identified epigenetic biomarkers of age in the Short-tailed shearwater (*Ardenna tenuirostris*) (De Paoli-Iseppi *et al.* 2017b). We aimed to create a rapid and relatively high-throughput epigenetic age assay using seven biomarkers previously identified using digital restriction enzyme analysis of methylation (DREAM) (Jelinek & Madzo 2016). We showed that these biomarkers could estimate age with an error of approximately three years. Here, we present DNA methylation (DNAm) age estimates for known- and unknown-age cohorts using targeted next generation sequencing (NGS, Illumina MiSeq platform).

5.3 Methods

5.3.1 Samples

For biomarker validation using NGS, bisulphite converted DNA samples (N = 190) were included in a single MiSeq assay targeting seven aDMPs identified in our previous research (Table 1). This run included a larger cohort of known-age samples (N = 109, including chicks), minimum-age individuals (i.e. banded as adults, N = 55) and unknown-age individuals (N = 24, 1 failed NGS). Known- and minimum-age individuals were collected from Fisher Island, Tasmania (40°13'00.7"S 148°14'20.7"E), and unknown-age birds from Little Dog Island, Tasmania (40°15'20.6"S 148°12'42.9"E). For minimum-age individuals, the time between initial banding and subsequent recapture is known. Samples were also grouped into age classes for some analyses as follows, chicks (< 1 year), young breeders (5 – 9 years), middle (10 – 18 years) and old (19 years or older). Sexes included in each group are shown in Supplementary Table 1.

5.3.2 Multiplexed restriction-site PCR (mRS-PCR)

Known DNA sequences (77 bp) from previous DREAM data were used to design specific reverse primers (R1 and R2, Supplementary Table 2) that were multiplexed with four universal, restriction enzyme recognition site based forward primers. The recognition motifs of four restriction sites were used, *Bam*HI, *Taq*I, *Sau*3A and *Eco*R1. These forward primers were combined with a T7 sequencing primer and ten degenerate (N) bases (Supplementary Table 2). Briefly, the first round was a 20 µL mRS-PCR using 2 µM of SRP1, 20 µM of each of the four restriction site oligonucleotides (RSOs), 10 µL of AmpliTaq Gold, 50 – 100 ng of purified blood DNA and nuclease-free water (Sarkar *et al.* 1993; Weber *et al.* 1998). For the second round, or nested PCR, 1 µL of the first-round product was used as a template with SRP2 and all other conditions the same. A total of 5 µL of the nested mRS-PCR was run on a 2% agarose gel using GelGreen® Nucleic Acid Gel Stain. First and second round cycling

conditions are shown in Supplementary Table 3. Nested PCR product was then cleaned using AMPure XP beads (1.8x) and used as a template for a sequencing reaction. Genomic sequencing was done in a 10 µL reaction using 0.25 µL of BRDT, 1.85 µL BigDye™ Terminator 5x sequencing buffer, 1 µL primer (3.2 nM), nuclease-free water and 2 µL of mRS-PCR template. Sequencing reactions were cleaned with CleanSEQ® before being sequenced using an ABI3130. If necessary, a band stab (2% agarose gel) of the mRS-PCR product was used as a template for an additional 20x cycling step, and this product was prepared for sequencing as above (Supplementary Table 4).

5.3.3 Bisulphite converted PCR (bsc-PCR) and library preparation

Upstream genomic sequence generated by mRS-PCR was used to create shearwater specific DNA templates to design bsc-PCR primer pairs. Isolated blood DNA (50 ng – 1 µg) was bisulphite converted for DNA methylation analysis using a Zymo EZ DNA Methylation-Lightning Kit (D5030) following the manufacturers' instructions. Primers were designed using MethPrimer (Li & Dahiya 2002) and did not contain any CG pairs within the primer sequences. Bisulphite converted primer pairs for *ELOVL2* and *KCNC3* were also included to test confirm previous observation with a larger set of known-age birds (De Paoli-Iseppi *et al.* 2017b). Bisulphite specific primers were temperature optimised (45.0 °C – 62.0 °C) and run with a no template control (NTC) and genomic DNA control to ensure specific amplification. All primer optimization was visualised using pre-stain gel electrophoresis (1.5% agarose, 100 V, 50 mins) with GelGreen® Nucleic Acid Gel Stain (Biotium, 1:10000). Cycling conditions for bsc-PCR are shown in Supplementary Table 5.

Bisulphite library preparation for use on the Illumina MiSeq platform was done as previously described (De Paoli-Iseppi *et al.* 2017b). Briefly, bisulphite-converted blood DNA was amplified and post-PCR products were diluted 1:10. To uniquely identify samples, i7 (N = 20) and i5 (N = 48) barcodes were added to the end of MiSeq universal primers. Individual plates were pooled and cleaned using AMPure (1.8x). Barcode-tagging efficiency was checked using a Bioanalyzer. All samples were then pooled into a single 1.5 mL Eppendorf tube and diluted to 4 nM. The library was prepared for sequencing on an Illumina MiSeq platform following the manufacturer's instructions with 20% PhiX spike-in sequencing control.

5.3.4 Data analysis

Demultiplexed fastq files were converted to fasta and Phred (Q) scores were checked using USEARCH v10 (Edgar 2010). A maximum expected error (maxee) of 0.7 was used to filter

poor quality reads. Targeted age-related differentially methylated positions (aDMPs) within 'CCCGGG' restriction cut sites were identified in amplicons using a script that used a unique 'core' sequence approximately five base pairs upstream of each CpG site identified prior to analysis (De Paoli-Iseppi *et al.* 2017b). Additional CpG sites within the new reference amplicon were also checked for age-related DNAm changes. The methylation level of each CpG was calculated as the number of methylated cytosines divided by the total number of reads for each site and recorded between 0 and 1, where 0 is unmethylated and 1 is methylated. A read depth of 100 was required to include a score for further analysis. Scores calculated using a read depth below this cut-off were discarded. Linear regression was used to assess the relationship between known age and DNAm level in technical replicates to validate scores within this run and observations from our previous DREAM data (De Paoli-Iseppi *et al.* 2018). Differentially methylated positions were included in a multiple linear regression model based upon Akaike Information Criterion (AIC).

Table 1 Next generation sequencing validation sample details

Method	Number of samples	Technical replicates (TR)	Known age *(N)*	Chicks *(N)*	Minimum age[#] *(N)*	Unknown age *(N)*	Known age range (years)*
Bisulphite converted NGS (MiSeq)	190 (+2 NTC)	15^	76	20	55	24	0.15, (5 - 21)

* Known-age represents ages of technical replicates.

^ STS16080B failed to sequence, so this TR was removed.

[#] Minimum age refers to birds first banded as adults.

Table 2 Bisulphite converted primers for target gene amplification.

Restriction site / biomarker	Direction	Sequence	Tm (°C)	bp	Product (bp)
M1934_bs_F2	F	AAGATGATTTATATATTTTGTTTTAGTAGT	53.2	30	83
M1934_bs_R1	R	TTAAAATTTCATTACTACTTATTCAATAAT	53.8	30	
M2083_bs_F1	F	GGATTTTTTTAGGGTTGAGTTTTTT	59.0	25	155
M2083_bs_R1	R	CCTTAAAAATCCTACCTTCCTCCT	59.0	24	
M3784_bs_F1	F	TTGGTTAAGTTGTTAGGAAGGAAAT	58.6	25	222
M3784_bs_R1	R	TCAACATCTCTCAACCCAAATATAC	58.5	25	

5.4 Results

5.4.1 mRS-PCR upstream genomic sequencing

Multiplexed restriction-site PCR of previously identified DREAM biomarkers was carried out to identify age-related methylation in CpG sites and obtain genomic sequences for PCR amplification and downstream analysis using the MiSeq platform. We amplified and sequenced five biomarkers from the original seven described in our previous study (De Paoli-Iseppi *et al.* 2018). The mean upstream fragment size amplified using this method was 306 bp (range: 168 – 518 bp). These genomic sequences were then used as input to BLASTn and 4/7 had a relatively high quality hit to an existing genomic entry (Supplementary Table 6). Example gel electrophoresis of first and second round nested mRS-PCRs is shown in Supplementary Figures 1 and 2, respectively.

5.4.2 High-throughput bisulphite converted NGS (MiSeq)

Using genomic sequence information generated by mRS-PCR, we designed primers specific for bisulphite converted DNA in our target amplicons. Three of these primer pairs (M1934, M2083 and M3784) successfully amplified the correct target (Table 2). The remaining markers either failed to generate a bsc-PCR product or had relatively high CG content, limiting the design of bisulphite converted primers. Amplicons were also included for *ELOVL2* and *KCNC3* to further explore DNAm in these genes with a larger data set. Bioanalyzer quantification also indicated high quality attachment of barcodes (Supplementary Figure 3).

Following sequencing using a MiSeq platform (Ramaciotti, Sydney), read depth levels were checked for each of the pooled amplicons in the library. Read depth was consistently high in known-age samples (mean range: 7743 – 12851 reads), with some expected variability between amplicons (Supplementary Figure 4A). Read depths in the minimum-age (Supplementary Figure 4B, mean range: 7735 – 12881 reads) and unknown-age samples (Supplementary Figure 4C, mean range: 7737 – 12136 reads) were also relatively high and consistent with known-age samples.

5.4.3 Conserved aDMPs, age-correlation in nearby CpG sites and amplicon variability

DNAm levels in aDMPs identified previously using the DREAM (CC**CG**GG) method were quantified, in addition to all other nearby CpG sites within each amplicon. We identified seven and eleven additional CpG sites for gene amplicons M2083 and M3784, respectively. One CpG site in M3784 identified from our Sanger sequence was not observed in the NGS data, resulting in a total of ten additional CpGs for the high-throughput run. No additional CpGs were noted in M1934. We quantified DNAm in eight and nine CpG sites in *ELOVL2* and *KCNC3* amplicons, respectively.

The DNAm levels in all eight *ELOVL2* CpG sites did not show any age association with the samples in our known-age set (Supplementary Figure 5A – H). CpG sites in *ELOVL2* were not used for any further analysis. DNAm levels in *KCNC3* also did not show any correlations with age (Supplementary Figure 6A – I). Age associations in aDMPs previously identified using DREAM, M2083-7 (Supplementary Figure 7F, $R^2 = 0.451$), M3784-11 (Supplementary Figure 8J, $R^2 = 0.185$) and M1934-1 (Supplementary Figure 9, $R^2 = 0.651$), were conserved in this NGS run. Linear regressions for all nearby CpG sites in M2083 and M3784 amplicons are shown in Supplementary Figures 7A – H and 8A – K respectively. We also observed similar age relationships for CpGs in close proximity to our previously identified aDMPs. In amplicon M2083, seven nearby CpG sites all showed a change in DNAm with known-age. In amplicon M3784, an additional eight CpG sites also showed DNAm changes with known-age. All CpGs analysed in this run and associated *P* and R^2 values are shown in Supplementary Table 7.

DNAm levels were also variable between each of these amplicons. DNAm levels remained relatively high and consistent across the entirety of the *ELOVL2* amplicon (Figure 1A). In contrast, both M2083 and M3784 amplicons show relatively large changes in DNAm. In M2083, a decrease in DNAm is observed from CpG-3 (mean = 0.811) to CpG-8 (mean = 0.179, Figure 1B). In M3784, the opposite relationship is seen, with an increase in DNAm from CpG-3 (mean = 0.039) to CpG-11 (mean = 0.656, Figure 1C).

5.4.4 Testing an age prediction model for known-age Short-tailed shearwater

DNAm data from a total of four aDMPs were used to create a Shearwater epigenetic age assay (SEAA). Separate linear regressions were done for each of the four aDMPs to investigate any sex-related DNAm effects. DNAm did not show any changes with sex in

these selected aDMPs (Supplementary Figure 10A – D). We chose to include chick samples and use aDMPs M2083-4, -5, -8 and M1934-1 in the model, where DNAm levels for M2083-8 and M1934-1 were log transformed (Figure 2, $R^2 = 0.728$, y intercept = 2.764 years). This four-aDMP assay provided DNAm age estimates, in known-age samples, with an overall mean absolute difference (MAD) of 2.71 years ($N = 109$, ± 2.02 years). The y-intercept of 2.764 years indicated that DNAm ages were overestimated for chicks and young birds. Model diagnostics and Q-Q plot suggest that the data is normally distributed as no Studentised residuals had a Bonferroni corrected $P < 0.05$ (Supplementary Figure 11A – B). The known-age cohort overall MAD values ranged from 0.057 – 9.7 years (Figure 3A). MADs for each group, as described in the methods, were lower in chick samples and decreased in middle-aged animals compared with the other groups (Figure 3B). DNAm level standard deviation between within-run technical replicates ($N = 14$) was relatively low with a mean difference of 4.9% (Supplementary Figure 12, Supplementary Table 8, range: 10.4 – 1.2%).

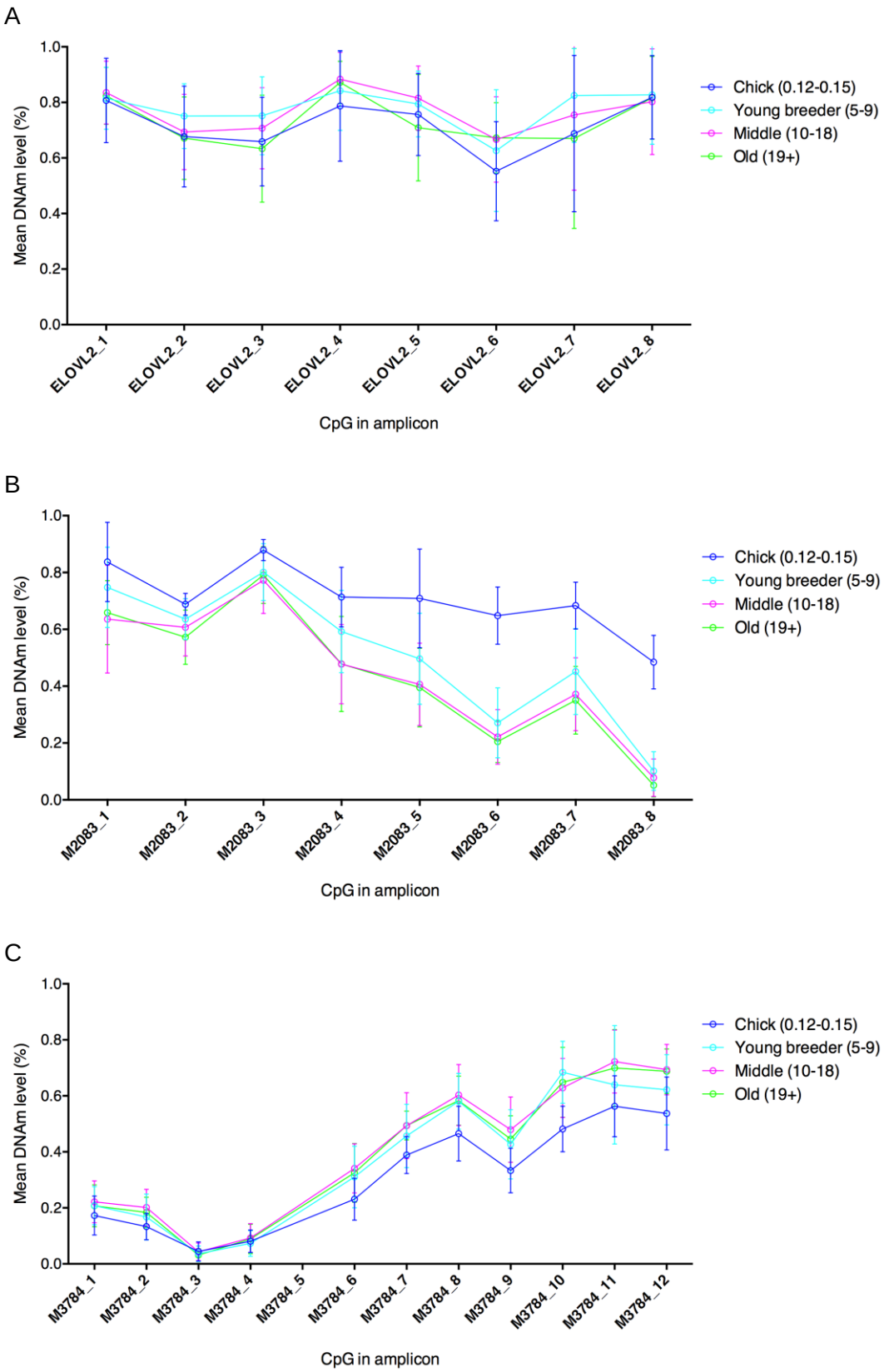


Figure 1A – C. Mean amplicon-wide DNAm levels. The mean DNAm level per CpG site (5' to 3') for each age group is shown for **A.** *ELOVL2*, mean amplicon levels in this amplicon are relatively consistent. **B.** Amplicon M2083 shows hypomethylation from CpG-3 onwards. The strongest aDMPs are within these hypomethylated CpG sites. **C.** In contrast to M2083, hypermethylation in the 5' to 3' direction is observed in the M3784 amplicon. The strongest aDMPs follow a drop to almost nil DNAm levels in CpG-3. Error bars indicate the standard deviation for each group.

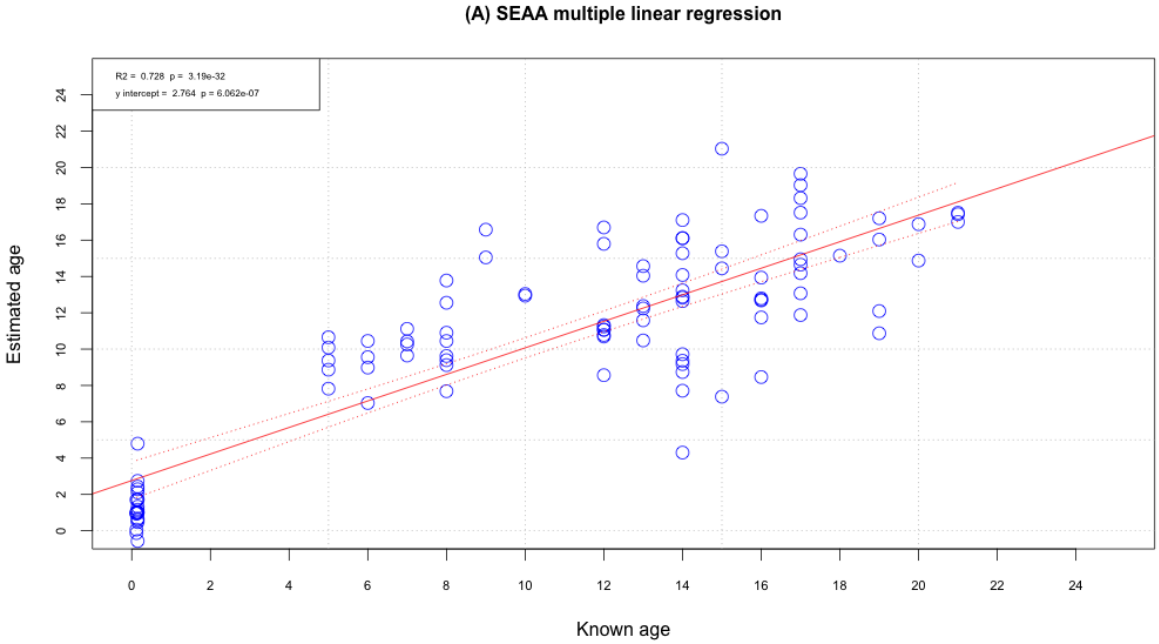


Figure 2. Full multiple linear regression model. Multiple linear regressions for estimated DNAm ages of *N* = 109 known-age Short-tailed shearwater from four CpG sites. 95% confidence limits of the placement of the regression line are shown in red.

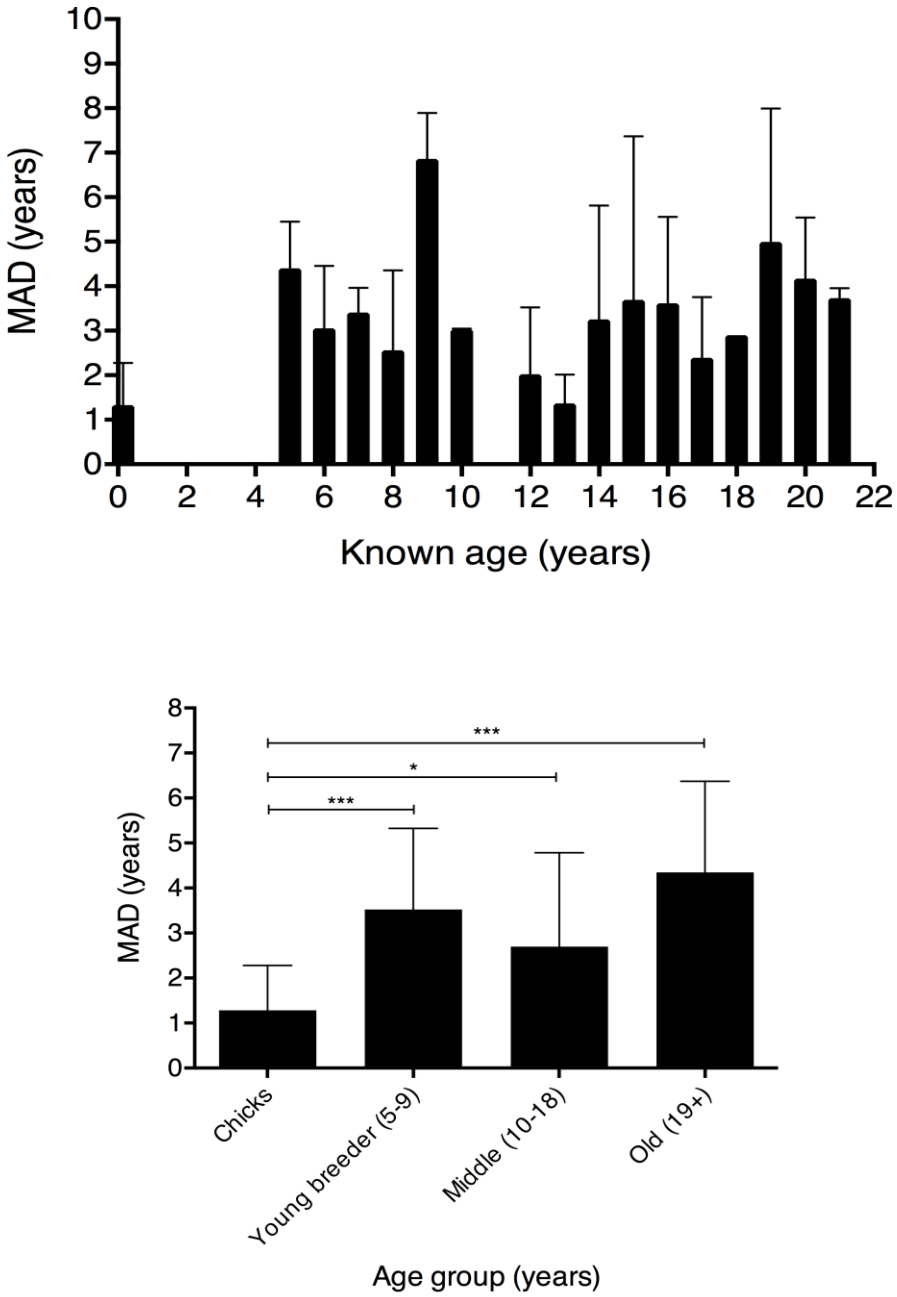


Figure 3A – B. Yearly and age-class grouped mean absolute deviation (MAD) for known-age samples. Determined by the absolute difference between the estimated and known age, the MADs are shown **A.** individually for each year of age for known-age animals included in the model and **B.** for each of the described age groups. Note: * $P < 0.05$ and *** $P < 0.001$.

5.4.5 Estimating DNAm age for a minimum-age cohort

We estimated the DNAm ages of minimum-age individuals (*N* = 55). These birds were banded as adults and subsequently resighted. The number of years elapsed between banding and recapture is the minimum age. We found that 71% (*N* = 55) of samples had DNAm age estimates greater than the known minimum-age. As there is some uncertainty around these ages, we also analysed the data given two ages for each individual. These were the unadjusted minimum-age and the minimum-age plus four years (approximate return date of young breeders) (Bradley *et al.* 1991; Bradley *et al.* 2000b). Estimates of DNAm age in the minimum-age cohort, using the four-aDMP model showed a trend with minimum age plus four years (R^2 = 0.141, Figure 4). Overall MAD was lowest for all minimum-age samples plus four years (3.71 ± 2.48 years, Supplementary Figure 13). The unaltered minimum-age had a MAD of 4.38 ± 2.89 years. MADs for each of these minimum age categories grouped by descriptive age-class show a decrease in MAD for middle age samples.

5.4.6 Estimating DNAm age for unknown age samples and population age structure

DNAm age was then estimated for unknown age samples using the model described above. The mean estimated DNAm age was 13.71 years (Figure 5, *N* = 23, range: 6.7 – 20.7 years). The known-age sample cohort (*N* = 86, with technical replicates and excluding chicks), showed an age-class composition with 26.7% of samples categorised as young breeders. A relatively larger proportion was middle aged (62.8%), whilst relatively few captured birds were classed as old (10.5%). The estimated DNAm ages for this sample set accurately reflected this known composition (Figure 6). However, fewer samples were allocated into the old class (3.5%). The DNAm age estimates for the minimum- and unknown-age animals were then combined into a single group (*N* = 78), given the uncertainty of individual provenance. This combined group had a structure comprised of young breeders (11.5%), middle (82.1%) and old individuals (6.4%, Figure 6).

5.4.7 Longitudinal observations of DNAm in resampled individuals

Several known- and minimum-age individuals were recaptured over a one to two year period. A total of 16 (76.1%) known-age samples (*N* = 21 pairs), showed the expected increase in DNAm estimated age between recapture. Fifteen of these pairs were recaptured after one year, with the remaining six recaptured after two years (Supplementary Figure 14A). In contrast, seven (43.8%) minimum-age samples (*N* = 16 pairs) showed the expected increase

in age between recaptures estimated by DNAm. Thirteen of these pairs were recaptured after one year with the remaining three recaptured after two years (Supplementary Figure 14B).

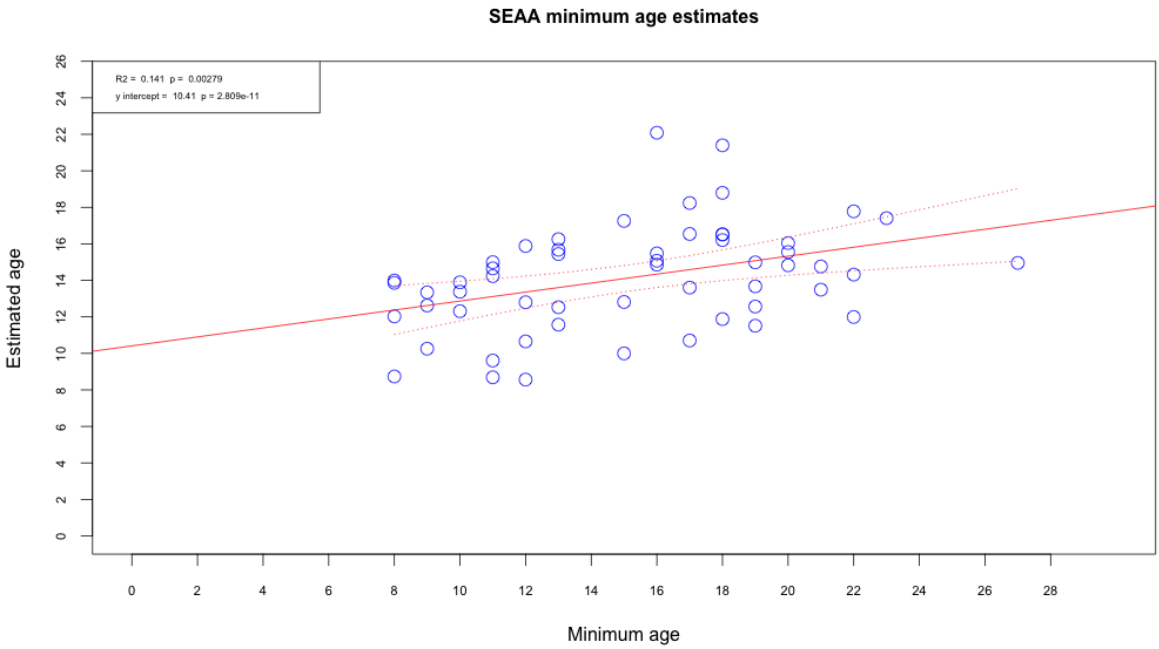


Figure 4. DNAm age estimates of minimum age samples. Linear regression for estimated DNAm ages of *N* = 55 minimum-age Short-tailed shearwater from four CpG sites. The plotted minimum age (x-axis) is the original minimum age plus four years (+4 years). 95% confidence limits of the placement of the regression line are shown in red.

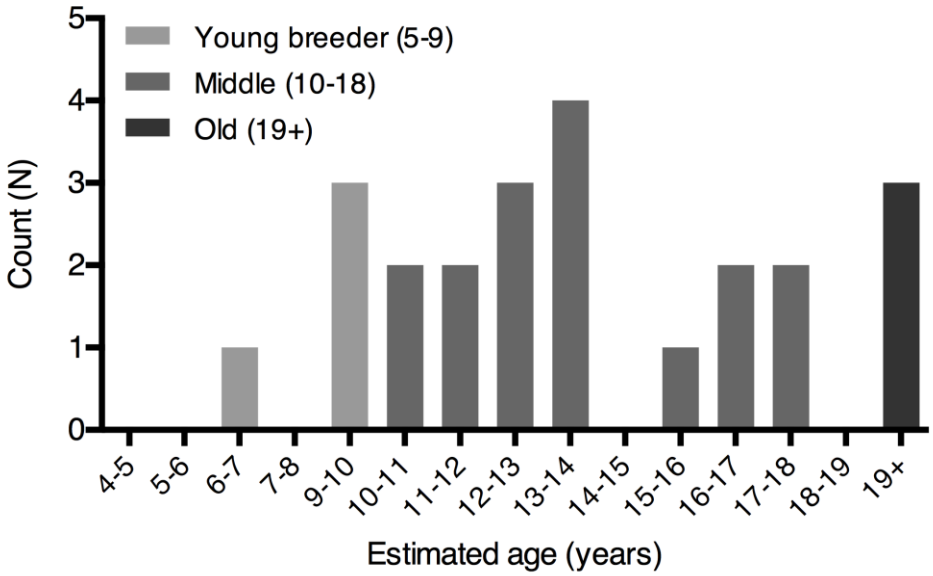


Figure 5. Unknown sample DNAm age estimate distribution. This distribution shows the count of individuals allocated to yearly age steps. These DNAm ages were estimated using methylation data generated using NGS (MiSeq) for a four-aDMP model. The markers used were M2083-4, -5, -8 and M1934-1. Grey shading indicates the age-class allocation.

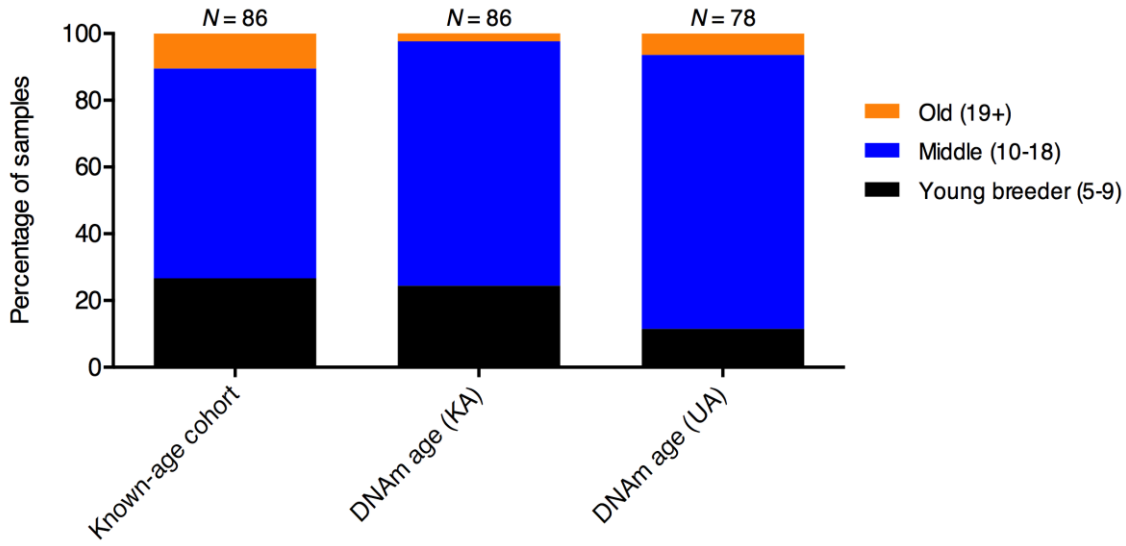


Figure 6. Known and estimated DNAm population age structures. The known population structure is shown for the banded, known-age cohort (*N* = 86, Fisher Island). The estimated population structure using DNAm is then shown for the known-age (KA) cohort and the minimum-age (MA) cohort (*N* = 55, Fisher Island) combined with the unknown-age (UA) cohort (*N* = 23, Little Dog Island). Young-breeders, middle age and old classifications are shown in black, blue and orange respectively.

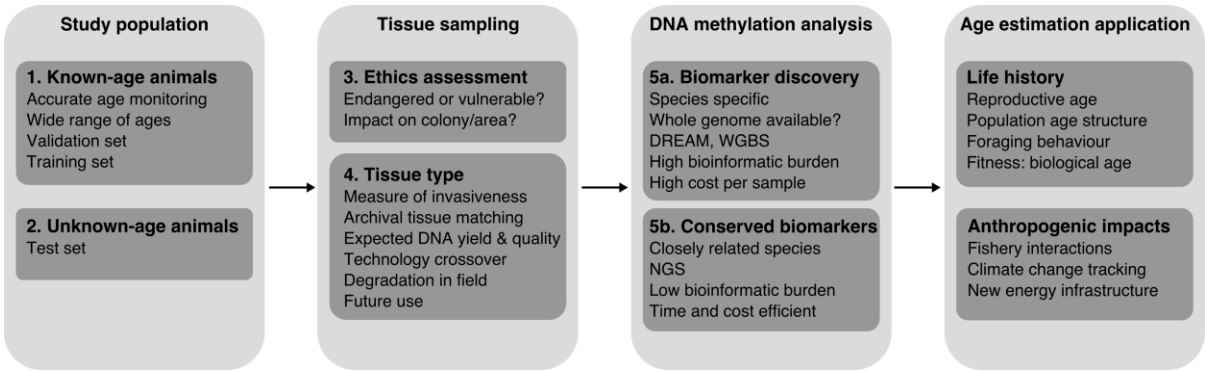


Figure 7. Considerations for non-model animal ageing studies using DNA methylation. Common considerations, pathways and goals for researchers wishing to investigate DNA methylation and ageing in wild animals. Abbreviations refer to digital restriction enzyme analysis of methylation (DREAM), whole genome bisulphite sequencing (WGBS) and next-generation sequencing (NGS).

5.5 Discussion

The age of many wild animals is difficult or impossible to estimate in cross-sectional population studies. Age distribution is an important aspect of the population biology of most species because it strongly affects population growth rates. There has consequently been significant effort invested in developing a range of age estimation methods that might inform population biology research on previously un-studied populations. In birds, attempts to estimate age using methods applicable to cross-sectional studies include quantification of pentosidine (Cooey *et al.* 2010; Dorr *et al.* 2017) and telomere length (Cerchiara *et al.* 2017; Tricola *et al.* 2018; Young *et al.* 2013). DNA methylation analysis is one of the most promising age estimation alternatives for a range of vertebrates (De Paoli-Iseppi *et al.* 2017a).

Previously, we showed that changes in DNAm with chronological age can be measured in the Short-tailed shearwater (*Ardenna tenuirostris*) (De Paoli-Iseppi *et al.* 2018). Specific DNAm changes correlated with age in humans and other species are now well established (Wagner 2017) and are being used to inform disease states or estimate an individual's biological age and general health (Horvath *et al.* 2014; Horvath & Raj 2018). Whilst this phenomenon is well studied in mammals, little research has been done in other vertebrates such as birds, reptiles, amphibian or fish (Ito *et al.* 2017; Polanowski *et al.* 2014; Wright *et al.* 2018). However, the broad conservation of this biomarker class, in addition to our recent research, indicates that DNAm age biomarkers may be useful in birds (De Paoli-Iseppi *et al.* 2017a). However, several considerations must be made before applying this aDMP method to new or closely related birds or reptiles. These considerations include the availability of a known-age validation cohort, tissue sampling impacts, DNAm analyses and the final application of age estimates (Figure 7). In this study, we quantified DNAm levels in target genes of known- and minimum-age Short-tailed shearwater. We also present the first DNAm age estimates for a cohort of unknown-age individuals.

5.5.1 A high-throughput method for screening age-related CpGs and estimating seabird age

Primers amplifying genes *ELOVL2* and *KCNC3* were the same as previously used (De Paoli-Iseppi *et al.* 2017b). With a relatively larger set of known-age individuals we confirmed our previous results that showed consistent DNAm levels with age in these genes. This is in contrast to findings in humans, particularly for *ELOVL2*, which has several age-related CpG sites in the promoter region (Slieker *et al.* 2018; Zbieć-Piekarska *et al.* 2015a). These results

could indicate that these specific regions in the shearwater are not regulated in the same manner as in humans or mice, and therefore markers of biological ageing in these genes are absent.

A limitation of markers identified using the DREAM method is that the 77 bp DNA sequence is too short to allow for simple design of genomic or bisulphite converted primers to amplify a CpG of interest. Multiplexed restriction-site PCR was a relatively effective method for acquisition of longer DNA sequences for this purpose, however four of the original seven aDMPs were not assayed due to amplification failures. BLAST assignments of reads to known genes were improved by obtaining longer DNA sequences. This demonstrates the value of this method for generating unknown sequence data for any animal species without an annotated whole genome.

Following mRS-PCR, we identified several additional CpG sites within the DNA sequences for M2083 and M3784. Interestingly, we also found that seven other CpGs in both M2083 and M3784 showed similar DNAm changes with chronological age. Hypo- and hypermethylation in additional CpGs for M2083 and M3784 respectively, was consistent with the direction of change in the original aDMP. These observations are not unusual, particularly for *ELOVL2*, which has been reported several times as having between seven to ten age-related CpG sites within the promoter region (Garagnani *et al.* 2012; Zbieć-Piekarska *et al.* 2015a). Four CpG sites within *EDARADD* have also been shown to be age-related in human saliva (Hamano *et al.* 2017). The multi-CpG age relationships observed for some specific genes are also conserved in other mammals and across tissue types including teeth, supporting our observations and suggesting that similar regions or genes can exist in seabirds (Polanowski *et al.* 2014; Slieker *et al.* 2018). Given that the DREAM method will only detect CpG sites within the 'CCCGGG' motif, it may be beneficial to examine other restriction enzyme combinations that could allow deeper analysis of the DNA methylome leading to incorporation of more CpG sites into our model increasing accuracy.

Our model with four aDMPs explained approximately 73% of variance with an overall MAD of 2.7 years. Whilst this was not as accurate as human or humpback whale epigenetic age assays, this method could be useful for quantifying large sample sets and quickly generating population snapshots (Horvath 2013; Polanowski *et al.* 2014). Chicks were included in the model as they further support the overall hypothesis that DNAm varies over the entire lifespan of the bird. In some aDMPs, M2083-6, -8 and M1934-1, an almost step-like change

in DNAm is noted between chicks and adults animals, possibly indicating a role for early developmental gene regulation. Whilst the overall MAD was relatively low, there was a difference between the error for chicks and all other age-classes.

In humans, cytosine methylation is known to influence transcription factor binding and the relatively large DNAm changes observed in both M2083 and M3784 could indicate the presence of a regulatory boundary within a gene promoter sequence (Yin *et al.* 2017). DNAm can be involved in gene repression and this may be what is occurring in M3784, as mean DNAm increases from approximately zero to 60 – 70% over a relatively short distance (Bird 2002). In M2083, there was a general decline in DNAm, which could be a signal of global hypomethylation known to occur in other species (Gryzinska *et al.* 2013; Nilsen *et al.* 2016). M2083 also had a partial BLAST hit to *TMBIM1*, which in humans negatively regulates matrix metalloproteinase 9 (*MMP9*) (Liu 2017). Proteins generated by members of the *MMP* family are involved in breakdown of the extracellular matrix and diseases including arthritis (Gruber *et al.* 1996). The enzyme encoded by this gene can also break down type IV and V collagens, possibly indicating a link to previous seabird ageing research that focuses on a marker of cross-linked amino acids in collagen (Chaney Jr *et al.* 2003; Fallon *et al.* 2006). M1934 had a partial hit to *RUNX1*, the up-regulation of which has been associated with aged stem cells in humans. The authors speculate that the increased expression of this gene with old age may facilitate leukemic transformation (Rossi *et al.* 2005). However, functional studies of gene expression would be required to determine if the observed DNAm has any impact on regulation.

The minimum-age cohort functions as a test set for our SEAA. Due to the uncertainty in return time of individuals to their natal island or colony, we chose to compare age estimates against two possible ages for each individual. Our model explained 14% of the variance and the MAD between the plus four-year adjustment and DNAm age was the lowest at 3.7 years. This may indicate that young non-breeding or immigrant un-banded individuals return to the colony at approximately four years of age, which supports previous findings at Fisher Island (Bradley *et al.* 1991).

5.5.2 Generating population age structure estimates using DNAm

For the first time, we have estimated the DNAm age of unknown-age shearwaters. The blood samples from the unknown age birds were collected from a nearby colony (Little Dog Island, Tasmania), with no information on chronological age other than assumptions based on other banded shearwater colonies. Due to the uncertainty beyond minimum-age, that is the time elapsed between recapture, animals that have been banded as adults are essentially of unknown age. We showed using our novel age assay that we could estimate the age of un-banded birds and closely replicate the known population age structure. Whilst this is encouraging for the field, it is important to note that these age structures have been generated using relatively small sample sizes, and that further studies using larger sample sizes are necessary to validate this initial study. The overestimation in the number young-breeders in the minimum-age cohort is an additional limitation, as these individuals can occasionally be classified as middle-aged. These limitations could be reduced in further studies with the addition of samples from 1 – 4 and 19+ year-old individuals.

Several closely related species exhibit similar life-history traits to the Short-tailed shearwater and therefore may be suitable for use in a similar epigenetic age assay (De Magalhaes & Costa 2009). These include the Manx shearwater (*Puffinus puffinus*), Wedge-tailed shearwater (*Ardenna pacifica*) and Sooty shearwater (*Ardenna grisea*). These species vary in their maximum wild longevity, approximately 51, 29 and 34 years respectively (Brooke 2013; Harris 1966). However, the age at sexual maturity for females is similar to that in the Short-tailed shearwater at 4.9 and 4 years for Manx and Wedge-tailed shearwater, respectively (Burger *et al.* 2001; De Magalhaes & Costa 2009). Close parallels can be drawn between the Sooty and Short-tailed shearwater as both are harvested either commercially or by indigenous tradition. Previous studies have shown that there is selective removal of heavier chicks during harvesting. However as initial weight does not carry through to post-fledge age, any impact of this selection, and therefore possible cryptic traits, remains unclear (Hunter & Caswell 2005). Harvesting, paired with removal of adults through fisheries bycatch makes assessing the sustainability of this practice a complex task (Uhlmann 2003). DNAm age estimates could also be used to model age-class interactions within model frameworks, such as the spatial population abundance dynamics engine (SPADE), which aims to inform and assess the management and conservation of wild species (Beeton *et al.* 2015).

Chapter 6 – General Discussion and Future Directions

"Don't believe anything I've told you—merely because I said it."

Arthur C. Clarke

The Songs Of Distant Earth

6.1 Overview

Chronological age is an important factor in understanding animal population biology. It also influences a host of individual factors including body condition, reproductive output and success, foraging behaviour and senescence (Elliott *et al.* 2015; Froy *et al.* 2013). Colonies of long-lived seabirds are banded and followed throughout the world and can be used as indicators of aquatic ecosystem health (Mallory *et al.* 2010). The costs of ensuring that records for these remote colonies continue annually or even for a single generation, which in some cases can be up to 50 years, can be considerable (Mallory *et al.* 2018). Using the Short-tailed shearwater (*Ardenna tenuirostris*) as a model species, we used age-related differentially methylated positions (aDMPs) in the bird epigenome to estimate age in known- and unknown-age individuals.

This thesis contains studies that investigated the development, validation and application of these molecular age biomarkers (MABs) in the shearwater. Data chapters (Chapters 3 – 5) include a discussion of the results based upon the available literature at the time. This final chapter reviews the most recent DNA methylation (DNAm) ageing studies and concludes with a discussion on the potential future applications of such biomarkers.

6.2 Recent DNA methylation based ageing studies

Several age-related DNAm studies have been published over the last year (Table 1) and some of these have not been mentioned in some chapters (chapters 2 – 3) of this thesis as they were not yet available. These recent studies have primarily focused on DNAm age estimation in mammals and have used both targeted and global approaches. These three studies focused on ageing wild chimpanzees (Ito *et al.* 2018), bats (Wright *et al.* 2018) and wolves (Thompson *et al.* 2017). Only one new study investigated DNAm in grouse (Soulsbury *et al.* 2018) and all of these will be discussed further here.

Thompson et al. (2017) used RRBS to collect DNAm data for 62 Gray wolves (*Canis lupus*) in order to investigate any age-related CpGs. As is the goal for many studies of DNAm in wild species, the authors sought to provide alternatives to existing ageing methods and their limitations. A total of 115 aDMPs were combined from both dogs and wolves to generate a DNAm age relationship ($R^2 = 0.8$). A relatively high proportion of these aDMPS are also aDMPs in humans. The authors note some limitations, including a low sample size ($N = 108$)

and some inaccuracy around the known ages of wild wolves that may affect the initial training set model (Thompson *et al.* 2017).

A study of wild Bechstein's bats (*Myotis bechsteinii*) quantified DNAm for aDMPs previously identified in mammals, an approach similar to the one used in our study (De Paoli-Iseppi *et al.* 2017b). DNA was extracted from the wing punches of 60 female, known-age and two deceased juvenile bats. The authors identified seven CpG sites within *TET2*, *GRIA2* and *ASPA* that predicted chronological age with an error of 1.52 years ($R^2 = 0.58$). Interestingly, the DNA extracted from wing punches here could be viewed as equivalent to previous age-related integument results in humans (Grönniger *et al.* 2010) and whales (Polanowski *et al.* 2014). Finally, eleven aDMPs in *ELOVL2* and *CCDC102B* were assessed in wild chimpanzees (*Pan troglodytes*) (Ito *et al.* 2018). The authors reported associations with age only within *ELOVL2*, which has previously been posited as a potential forensic marker in humans (Zbieć-Piekarska *et al.* 2015a; Zbieć-Piekarska *et al.* 2015b).

These studies provide further evidence that existing mammalian aDMPs can be used to age wild species. However, despite these findings there is little DNAm research in birds (Head 2014; Jarman *et al.* 2015). In Black grouse (*Lyrurus tetrix*), DNAm correlation with age was recently reported (Soulsbury *et al.* 2018). In this study, a nonlinear correlation was identified in *AgRP*, which is involved in the production of pigments for sexual ornamentation in some birds. Sexually selected traits can show these responses with age and the authors note that this data will allow for further exploration of DNAm and individual condition in grouse. These markers may be useful for some seabirds species such as the Whiskered auklet (*Aethia pygmaea*) or the Atlantic puffin (*Fratercula arctica*), which have sexually selected plumage and pigmented beaks respectively (Harris 2014; Seneviratne & Jones 2008). However, genes associated with sexually selected ornamentation may not be good candidates for application in seabirds including shearwater species, that do not show similar physiological changes during the breeding season.

Table 1 Recent DNA methylation ageing studies in model and wild animals

Class/Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	MAD/error	Age relationship
Mammalia								
Chiroptera	*Myotis bechsteinii*	Wright et al. (2018)	Wing tissue	0 - 14 (years)	*TET2, GRIA2, ASPA*	*N* = 7	1.52 (years)	0.64
Primates	*Pan troglodytes*	Ito et al. (2018)	Blood	2 - 35 (years)	*ELOVL2, CCDC102B*	*N* = 11	5.41 (years)	0.741
Rodentia	*Mus musculus*	Meer et al. (2018)	Multi-tissue	0.2 - 35 (months)	-	*N* = 435	2.14 - 4.66 (weeks)	0.89
	Mus musculus	Thompson et al. (2017)	Multi-tissue	6, 12, 18 (months)	-	4 novel clocks. All CpGs: *N* = 193651, conserved: *N* = 952	2.5 - 3.8 (months)	0.68 - 0.82
Carnivora	*Canis familiaris*	Thompson et al. (2017)	Blood	0.5 - 14 (years)	-	Combined: *N* = 115	Combined: 0.8 (years)	0.8
	Canis lupus	Thompson et al. (2018)	Blood	0.5 - 8 (years)	-			
Aves								
Galliformes	*Lyrurus tetrix*	Soulsbury et al. (2018)	Blood	1 - 6 (years)	*AgRP*	CpG sites 2 - 5	-	Nonlinear (inverse u-shaped)

6.3 General discussion

Over the course of this thesis, we have investigated several methods of estimating age in the Short-tailed shearwater. In general, interest is increasing in robust and cost-effective methods that can be applied to wild animals. In humans and model organisms such as rats and mice, estimation of biological age is of great interest, and research is beginning to reveal associations between accelerated biological ageing and disease states (Horvath *et al.* 2014). The establishment of a known-age dataset for the purpose of marker discover however is generally, far simpler in humans than in wild species. Known-age datasets for wild animals are scarce, and require extensive levels of physical and financial expenditure to create even a small set. Ideally, a rapid biomarker of age for a target species would be deployed to determine age structure in unknown populations. In seabirds, few markers of age exist beyond fledging. In large unbanded colonies, no knowledge on age is currently available other than assumptions made from closely followed colonies or related species.

We have shown that most DNAm-age biomarkers in birds are likely to be different to those in mammals. Through targeted NGS we showed that markers commonly associated with age in mammals were not conserved in the shearwater. This led us to use a relatively novel global DNAm analysis method, DREAM, to assay thousands of CpG sites for age-related changes (Jelinek & Madzo 2016). Interestingly, this method has led us to find age-related differentially methylated regions (aDMRs) in the shearwater. This overall phenomenon appears to be conserved between mammals and birds, but the specific genes that are affected by age-related DNAm are different. The studies presented in this thesis are the first of their kind for a wild bird and have widened the possibilities for using DNAm in a conservation or management context. As such, our understanding of these biomarkers would be improved by the collection of larger known-age cohorts across a wider range of ages, particularly for generating age structures of unbanded colonies.

6.4 Questions and pathways for future studies

Further studies focusing on ageing seabirds could take several pathways depending on the hypothesis under consideration or overall project goal. Here, we focus on two broad areas, technical and ecological, and the advantages and limitations of both that could be used to further advance this field of study (Table 2).

6.4.1 Technical considerations and challenges

6.4.1.1 Alternative tissue sources

It is important to try and reduce the impact of scientists on wild animals. In the majority of this thesis, we have primarily taken blood samples from the foot webbing of shearwater. Whilst a relatively small volume was extracted, handling times of individuals are extended to allow for this step. Therefore, it may be beneficial to begin examination of alternative tissue sources of DNA including skin, or by proxy feather samples, flesh and faeces. However, one benefit of blood sampling is that birds have nucleated red blood cells, which allows for repeated testing of high quality DNA and associated long-term storage. If the same aDMPs can be identified in these alternate tissues it could allow for reduced handling times and stress in seabirds. In humans, several tissue types have now been used for age estimates, and in some genes, DNAm changes consistently across these tissues (Bekaert *et al.* 2015; Grönniger *et al.* 2010). If aDMPs or aDMRs were conserved across seabird tissues such as skin, faeces or buccal cells these could provide alternate and less invasive options for seabirds, particularly for those that are threatened. However, careful experimental design would be required to ensure that enough material is collected to ensure adequate recovery of high quality DNA. Tissues that do not yield a large amount of DNA, or from which DNA is degraded, can have downstream hidden costs and may be better suited to studies that use well-tested biomarkers (Bosnjak *et al.* 2013; McDonald & Griffith 2011).

6.4.1.2 Restricted global DNA methylation analyses

The DREAM discovery method used in chapters 4 and 5 of this thesis reports DNAm results for CpG sites within a ‘CCCGGG’ motif (De Paoli-Iseppi *et al.* 2018; Jelinek & Madzo 2016). This is due to the use of two restriction enzymes *SmaI* and *XmaI*, which both recognise, and distinctly cut, this specific sequence. This technique is broadly similar to reduced representation bisulphite sequencing (RRBS), in which the *MspI* enzyme recognises the ‘CCGG’ motif (Meissner *et al.* 2005). This four-base pair sequence will occur more frequently in the genome, and therefore more CpG sites can be quantified. The trade off when using this method is that a higher read depth will be required to ensure high confidence in DNAm scores. In chapter 5, we showed that read depth between separate DREAM runs may not be consistent, which limits our ability to confidently score the same aDMPs. It is also critical that when using these techniques that technical replicates are used, both within and between sequencing runs, so that investigators can report any variability in DNAm scores

(Robasky *et al.* 2014). For example, if the reported error rate or noise is equal to the range of age-related DNAm for a biomarker it may not be considered robust. Care must be taken to try and use the same DNA for these comparisons, as separate extractions will include different cells that can contribute to noise.

6.4.1.3 Whole genome and epigenome sequencing

Whole genome sequencing (WGS) followed by whole genome DNAm analysis is an additional pathway when starting with a species with no prior reference sequence. Whilst the costs of WGS has rapidly decreased, financial cost will be relatively high compared to restricted, higher throughput methods, as this technique would require a sufficient number of individuals representing the lifespan of the species (Wetterstrand 2013). The bioinformatics bottleneck, a common problem in several fields, would also increase given that *de-novo* assembly and functional annotation would be required (Escobar-Zepeda *et al.* 2015). Briefly, the bioinformatic pipeline would require *de-novo* genome assembly, followed by mapping of epigenomic markers. The advantage of this approach is that theoretically, the entire epigenome could be scanned for age-related changes and mapped back to a known gene. This approach provides the opportunity to considerably increase our basic understanding of age-related changes and examine conservation between animals. Declining costs of sequencing may make this approach cheap enough to be practical in the near future.

6.4.2 Ecological considerations and challenges

6.4.2.1 Known-age calibration and unbanded population cohorts

One challenge that is difficult to overcome when searching for aDMPs in wild animals is the large number of known-age individuals required to both train and test any candidate biomarkers. Our shearwater epigenetic age assay would benefit from a greater sample size and a wider range of ages that better represent the entire lifespan of the bird. This could be accomplished by increased field collections from individuals aged one to four and over 20 years. Due to variable behaviour and higher mortality in these young and old animals, a change in fieldwork protocols would be required. For example the tracking of these individuals through either live GPS or light geolocators data could reveal where these animals concentrate during non-breeding years or overwintering, allowing for more targeted sampling to fill in these gaps (Frederiksen *et al.* 2012; Louzao *et al.* 2009).

An increasing number of species, particularly birds, are being impacted by human behaviours including climate change (Pearce-Higgins *et al.* 2015). In order to provide accurate information for the conservation of these species, robust information on basic life history is essential. Therefore, the application of these novel aDMPs and aDMRs within actionable management or conservation scenarios could be very rewarding. Using DNAm age estimates for the construction of cohort life tables could be used to assess patterns in past mortality rates or make informed predictions of future mortality changes (Beissinger & Peery 2007). Life tables would be particularly useful for the Short-tailed shearwater, which is subject to an annual harvest. Several colonies on islands within the Furneaux Group, Tasmania are commercially and recreationally harvested. DNAm age estimates could be used to monitor population age structure between unharvested and variable pressure harvest colonies. However, as noted in a previous study, selective removal of chicks and therefore survival, from Sooty shearwater (*Ardenna grisea*) colonies had a relatively small impact on population growth (Hunter & Caswell 2005).

DNAm age estimates that can assign individuals into classes can also determine parent-offspring pairs, which is useful information as these birds are highly philopatric (Bravington *et al.* 2014; Polanowski *et al.* 2014). Our epigenetic age assay may be particularly useful within commercial fisheries to estimate individual age of seabird bycatch. Previous research shows that mortality of adult seabirds due to bycatch has a significant impact on population growth rate (Cortés *et al.* 2018; Hunter & Caswell 2005; Uhlmann 2003). If aDMPs can be identified in DNA extracted from flesh samples, this would make sample collection at sea comparably simple.

6.4.2.2 DNA methylation changes during early development in chicks

Shearwater chicks grow rapidly over the first three months post-hatching and can become twice the weight of their attending parents (Bradley *et al.* 2000a). This weight is then lost as the adults leave prior to chick fledging and most birds leave with limited fat deposits in reserve (Lill & Baldwin 1983). An experiment focused on sampling whole blood over this relatively rapid period of growth could allow further insight into DNAm changes during this time. Chicks undergo several changes including rapid growth to fasting state within weeks and DNAm plasticity may play a role in regulating developmental genes. Disentangling confounding environmental effects including low food availability, burrow flooding or high stress would be challenging (Bollati & Baccarelli 2010; Romano *et al.* 2017; Visser 2001).

Skin or feather aDMPs may also be associated with the shedding of downy feathers and growth of larger flight feathers. At this time, no studies have investigated DNAm changes with age over this time, and could reveal additional aDMPs that change into adulthood.

6.4.2.3 Conserved aDMPs in other seabirds and demographic traits

The shearwater used as the model species in this thesis have been historically monitored in one of the world's longest running banding studies (Bradley *et al.* 1991). As a model species that shares some life history traits with other Procellariformes, the data presented in these chapters provides a foundation for expansion into these species. Some of these demographic traits including the minimum-age of first reproduction and age-specific survival, are difficult to estimate as it requires recapture over several years (Weimerskirch 2001). These traits are also not likely to vary within populations, as they are less adapted to the local environment. However, breeding frequency and success, and age-class survival can be influenced by environmental conditions, which may also be responsible for the DNAm noise observed for individuals of the same age (Weimerskirch 2001). Our epigenetic age assay could be particularly useful for age structure analysis and detecting within-population survival variability due to differential environmental conditions, resource exploitation or anthropogenic impacts (Chambers *et al.* 2015; Clutton-Brock & Sheldon 2010; Gianuca *et al.* 2017; Lewison *et al.* 2012).

Examining the conservation of our aDMPs would follow the targeted NGS methods we outlined in chapter 5. Briefly, species-specific bisulphite-converted primers should be designed to amplify each aDMP in known-age individuals. Targeted design of primers will reduce the occurrence of single nucleotide polymorphisms between species that can reduce PCR efficiency. Age estimates generated through our assay could supplement existing long-term banding studies of vulnerable seabirds including the Wandering albatross (*Diomedea exulans*) (Fay *et al.* 2016; Wooller *et al.* 1992). DNAm age estimates could also be used to validate or compare ages determined using alternative methods including pentosidine or beak grooves (Harris 2014; Rattiste *et al.* 2015).

6.4.2.4 Other non-model wild species

In general, the techniques we used to identify shearwater aDMPs are advantageous for any animal that does not have a reference genome, as mapping target sites is not strictly necessary

to assess age relationships. Enrichment for gene promoter regions and CpG islands, often the target in human studies, can be accomplished by selecting for DNA fragments of approximately 40 – 200 base pairs (Gu *et al.* 2011). The estimation of DNAm age using non-lethal methods in fish would be a valuable method that could be integrated into existing studies that estimates abundance through tissue sampling (Bravington *et al.* 2016; Trenkel *et al.* 2019). There has been some interest in determining the age of reptiles, and there is some evidence to support global DNAm changes with age (Iverson *et al.* 2017; Nilsen *et al.* 2016; Wilson *et al.* 2003). However, given the evidence presented in this thesis, it is unlikely reptilian aDMPs will be exactly conserved with either mammals or birds.

6.5 Final thesis conclusions

Wild-animal epigenomics is rapidly generating interest thanks to several recent studies that have investigated age-related changes and decreasing costs associated with established technologies. An understanding of age-related variability in foraging, breeding success and survival is critical for population ecology. The research presented in this thesis supports the use of DNA methylation as a valuable technique for estimating age in the Short-tailed shearwater. The ability to generate population age structures from previously unbanded colonies will allow for expanded monitoring of seabirds. DNAm age estimates, when combined with other demographic traits, have the potential to detect historical and modern changes in seabird populations and have a meaningful impact on conservation and management.

Table 2 Pathways and improvements for future studies

Broad objective	Focus	Other information	Advantages	Limitations	Priority
Technical	Alternate tissue types	Skin (feather)	Medium to low invasiveness	Low DNA yield	Medium
		Faeces	Non-invasive	Low DNA yield & quality	Medium
		Flesh	High quality DNA, ideal for fisheries	Highly invasive for live animals	Medium
		Buccal cells	Low invasiveness	Low DNA yield	Low
	DREAM	Higher read depth and sequencing length	Greater reproducibility	Increased cost per sample	High
	RRBS	Quantify 'CCGG' DNAm	Greater coverage of CpG sites	Increased bioinformatic burden	High
	Whole genome sequence	Genomic sequence of target animal	Annotation and reference genes	High cost and bioinformatic burden	Low
	Whole epigenome sequence	Epigenomic sequencing	Quantify entire epigenome	High cost and bioinformatic burden	Low
Ecological	Known-age training set	Increase known-age samples, particularly at extreme ranges of lifespan	Increased understanding of age-correlations	High fieldwork costs	High
	Unbanded population age structures	Estimate DNAm age of large set of samples	Greater representation of colony	High fieldwork costs	High
	Chick biology	DNAm changes over the first months of life	Understand DNAm changes during early growth	Highly disruptive of nesting birds	Low
	Closely related seabirds	Investigate aDMP conservation	Application of epigenetic age assay to other long-lived seabirds	May require recalibration with large known-age training set	Medium
	Other wild species	aDMP conservation or discovery in reptiles or fish	Better understanding of age-related biomarkers in animals	Difficult to establish known-age datasets	Low

Supplementary Material

Chapter 2

Supplementary Table 1. Age-related DNA methylation changes in mammals, birds, reptiles and fish

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
Mammalia										
Primates	*Homo sapiens Avg. lifespan: 85 years**	Gronniger et al. (2010)	Skin (pilot study)	26 - 35 & 65 - 71 years	-	104	Pos			β-value + 0.2
			Skin (validation)	19 - 72 years	SEC31L2	1	Pos			p(BH) < 0.01
					DDAH2	1	Pos			p(BH) < 0.01
					TET2	1	Pos			p(BH) < 0.01
		Teschendorff et al. (2010)	Blood	52 - 78 years	-	69	Pos			p = 6e-15 (training, pre-treatment), p = 2e-06 (validation, postreatment)
		Bocklandt et al. (2011)	Saliva (model)	21 - 55 years	-	88	Neg = 19, Pos = 69			*q*-values < 0.05, pred vs obs (r = 0.83, p = 2.2x10-16)
			Saliva (validation)	18 - 70 years	EDARADD	1	Neg			r = -0.81 (twins)
					NPTX2	1	Pos			r = 0.52 (twins)
					TOM1L1	1	Neg			r = -0.70 (twins)
		Koch & Wagner et al. (2011)	Dermis, epidermis, cervical smear, blood. (Training set.)	16 - 72 years	-	5	na	10.3 years	12.11%	Model: N = 5 CpGs, pred vs obs (R^2 = 0.65)

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
			Saliva, blood (cord, peripheral), breast. (Validation set.)	0 - 78 years	-	5	na			Model: N = 5 CpGs, pred vs obs (R^2 = 0.68)
					TRIM58	1 (cg07533148)	Pos			r = 0.69 (all training set)
					KCNQ1DN	1 (cg01530101)	Pos			r = 0.63 (all training set)
					NPTX2	1 (cg12799895)	Pos			r = 0.62 (all training set)
					GRIA2	1 (cg25148589)	Pos			r = 0.62 (all training set)
					BIRC4BP	1 (cg23571875)	Neg			r = -0.45 (all training set)
		Garagnani et al. (2012)	Whole blood (training)	42 - 83 (mothers) & 9 - 52 years (offspring)	ELOVL2	1 (cg16867657)	Pos			r = 0.91, p = 3.97e-18
					FHL2	1 (cg06639320)	Pos			r = 0.91, p = 1.82e-18
					PENK	1 (cg16419235)	Pos			r = 0.76, p = 8.10e-15
			Whole & cord blood (validation)	9 - 99 years	ELOVL2	4	Pos			r = 0.92
					FHL2	4	Pos			r = 0.80
					PENK	2	Pos			r = 0.63
		Alisch et al. (2012)	Blood (test)	3 - 17 (mean = 9.9) years	-	2078	Neg = 1601, Pos = 477			FDR < 0.01
			Blood (validation)	1 - 16 (mean = 4.6) years	-	41895	Neg = 35997, Pos = 5898			FDR < 0.01, 62.7% overlap with test set (p < 0.001)
		Hannum et al. (2013)	Whole blood	19 - 101	-	71	na	3.9 years	4.58%	r = 0.963, ± 3.9 years (training), r = 0.905, ± 4.89 years (validation)

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
			Breast, kidney, lung, skin (multi-tissue validation)	-	-	71	na			r = 0.913, ± 5.71 years
		Horvath (2013)	Multi-tissue: e.g. blood, colon, adipose, liver, lung, saliva	Mean = 43 years	-	353	Neg = 160, Pos = 193	3.6 years	4.23%	r = 0.97, ± 2.9 years (test), r = 0.96, ± 3.6 years (validation)
		Weidner et al. (2014)	Blood	0 - 78 years	-	102	Neg = 58, Pos = 44	3.34 years	3.93%	R^2 = 0.98, ± 3.34 years (training), R^2 = 0.71 (Hannum (2013) validation)
			Blood (simplified model)	0 - 78 years	-	3	Neg = 2, Pos = 1	5.4 years	6.35%	Model: N = 3 CpGs, ± 5.4 years (training), ± 4.5 years (validation)
					ITGA2B	1 (cg25809905)	Neg			r < -0.85
					ASPA	1 (cg02228185)	Neg			r < -0.85
					PDE4C	1 (cg17861230 + 1)	Pos			r > 0.85
		Zbiec-Piekarska et al. (2015a)	Blood	2 - 75 years	ELOVL2	2 (Chr6:11044634, 11044642)	Pos = 2	5.75 years	6.76%	R^2 = 0.859, ± 5.03 years (training), ± 5.75 years (validation)
		Zbiec-Piekarska et al. (2015b)	Blood	2 - 75 years	-	5	na	3.9 years	4.58%	Model: N = 5 CpGs, R^2 = 0.94, ± 4.5 years (training), ± 3.9 years (validation)
					ELOVL2	1 (Chr6: 11044661)	Pos			p = 1.1121x10-23
					C1orf132	1 (Chr1: 207823681)	Neg			p = 3.1724x10-23
					TRIM59	1 (Chr3: 160450199)	na			p = 3.3826x10-12
					KLF14	1 (cg14361627)	na			p = 4.353x10-12
					FHL2	1 (Chr2: 105399288)	na			p = 2.3384x10-8

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
		Bekaert et al. (2015)	Blood	0 - 91 (mean = 44) years	-	4	Neg = 2, Pos = 2	4.96 years	5.83%	Model: N = 4 CpGs, pred vs obs (R = 0.97, p = 1.272e-129, ± 3.75 years, test). R^2 = 0.9469, ± 4.96 years, validation)
			Teeth	19 - 70 (mean = 39) years				4.86 years	5.71%	Model: N= 4 CpGs (R^2 = 0.74, ± 4.86 years)
					ASPA	1	Neg			CpG1: R^2 = 0.65, p = 1.4071e-07
					PDE4C	4	Pos			CpG1: R^2 = 0.85, p = 6.0245e-05
					ELOVL2	9	Pos			p = 1.2628e-11
					EDARADD	2	Neg			CpG1: R^2 = 0.62, p = 7.4543e-25
					ITGA2B	3	na			R^2 < 0.3 (not in model)
		Vidal-Bralo et al (2016)	Blood	20 - 78 years (training)	-	8	Pos	5.07 years	5.96%	R^2 = 0.68, ± 5.07 years
				45 - 89 years (MS-SNuPE validation)	-	8	Pos	6.07 years	7.14%	R^2 = 0.45, ± 6.07 years
		Christiansen et al. (2016)	Buffy coat	30 - 82 years (twins)	-	Hannum (71 CpGs), Horvath (353 CpGs)	na	5.6, 5.4 years	6.59%, 6.35%	Horvath (2013): r = 0.97, ± 5.6 years, mean = 1.4 years lower than chronological age. Hannum (2013): ± 5.4 years, mean = 3.6 years higher than chronological age

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
		Eipel et al. (2016)	Buccal cells	1 - 85 years	-	5	Neg = 3, Pos = 2			Model = "5 CpGs" (Buccal-Cell-Signature) pred vs obs, ± 4.66 years, R^2 = 0.93 (training), ± 5.09 years, R^2 = 0.93 (validation)
		Mawlood et al. (2016)	Blood	18 - 91 years	-	2	Neg = 2	9.3 years	10.9%	Model = 2 CpGs, pred vs obs: R^2 = 0.509, ± 9.3 years
					M1215	1	Neg			r = -0.322, p = 0.043
					M1313	1	Neg			r = -0.383, p = 0.015
	Pan troglodytes Max lifespan: 59.4 years	Horvath (2013), Pai et al. (2011)	Heart	~ 10 - 42 years (captive)	-	353	Neg = 160, Pos = 193	10 years	16.83%	r = 0.84, p = 0.00063, ± 10 years
			Kidney, liver	~ 9 - 46 years (captive)	-			3.7 years	6.23%	r = 0.75, p = 5.8e-05, ± 3.7 years
		Horvath (2013), Hernando-Herraez et al. (2013)	Blood	~ 9 - 34 years (wild born)	-	353	Neg = 160, Pos = 193			r = 0.93, p = 0.022
	Pan paniscus Max lifespan: 55 years	Horvath (2013), Hernando-Herraez et al. (2013)	Blood	~ 15 - 22 years (wild born)	-	353	Neg = 160, Pos = 193	1.4 years	2.54%	r = 0.84, p = 0.036, ± 1.4 years
	Gorilla gorilla Max lifespan: 60.1 years	Horvath (2013), Hernando-Herraez et al. (2013)	Blood	~ 16 - 44 years (wild born)	-	353	Neg = 160, Pos = 193	14 years	23.29%	r = 0.51, p = 0.3, ± 14 years
Rodentia	*Mus musculus*	Maegawa et al. (2010)	Small intestine	3 - 35 months	DOK5	na	Pos			r = 1.00, p < 0.001
					MYOD1	na	Pos			r = 1.00, p < 0.001
					CDKN2A	na	Pos			r = 0.96, p = 0.003

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
					IGF2	na	Pos			r = 0.86, p = 0.024
					NKX2	na	Pos			r = 0.86, p < 0.001
	C57BL/6J	Spiers et al. (2015)	Blood, lung, cerebellum and hippocampus	17 (embryonic) - 630 days	ELOVL2	2	Pos			p = 1.15e-04 (blood)
					GLRA1	2	Pos			p = 3.31e-06 (blood)
					MYOD1	1	Pos			p = 2.8e-06 (blood)
					PDE4C	2	Pos			p = 2.1e-06 (cerebellum)
	*C57BL/6-BABR Avg. lifespan >100 weeks**	Stubbs et al. (2017)	Multi-tissue: liver, lung, heart and brain	Newborn - 41 weeks old	-	329	na	3.3 weeks (training)	3.33%	r = 0.839 (training set)
	Rattus norvegicus	Penner et al. (2016)	Hippocampus	9 - 12 months (adult), 24 - 32 months (aged)	EGR1	20	Pos			Averaged % methylation of 20 promoter CpGs showed higher methylation in aged rats (p = 0.021)
Cetacea	*Megaptera novaengliae Max lifespan: 95 years*	Polanowski et al. (2014)	Skin	<1 - 30 years	-	3	Neg = 1, Pos = 2	3.75 years	3.95%	Model: N = 3 CpGs, pred vs obs (R^2 = 0.787, ± 2.991 years)
					TET2	1	Neg			R^2 = 0.455, p = 2.26e-07
					CDKN2A	1	Pos			R^2 = 0.471, p = 1.16e-07
					GRIA2	1	Pos			R^2 = 0.568, p = 1.38e-09

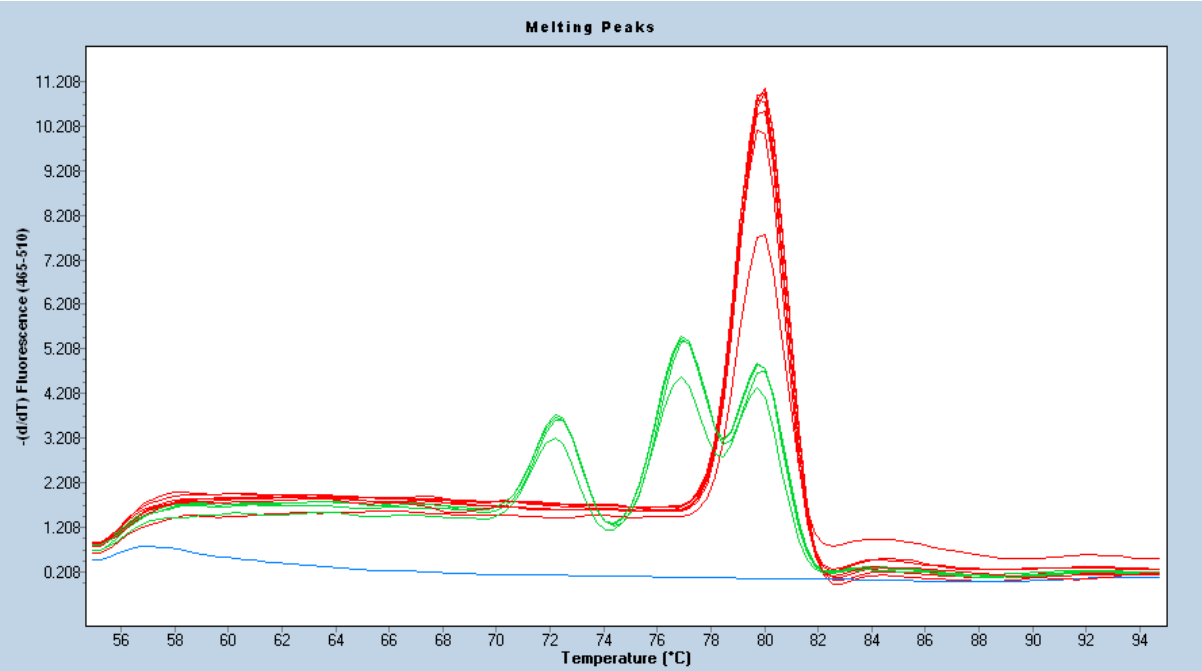
Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
Carnivora	*Canis lupus familiaris Max lifespan: 24 years*	Gryzinska et al. (2016)	Blood	2 months - 14 years	Global	-	Pos			Relative methylation level: 43.5% (<= 9 months) - 81.29% (> 8 years)
		Ito et al. (2017)	Blood	0 - 16.6 years	-	4 or 2	Neg = 2, Pos = 2	34.3, 23.1 months	11.88%, 8%	Model: N = 4 CpGs, MAD = 34.3 months, N = 2 CpGs, MAD = 23.1 months
					GSE1	1 (cg07082267)	Neg			R^2 = 0.236, p = 0.001
					SCGN	1 (cg06493994)	Pos			R^2 =0.066, p = 0.101
					BCL6B	1 (cg10137837)	Neg			R^2 = 0.049, p = 0.160
					POU4F2	1 (cg05991454)	Pos			R^2 = 0.136, p = 0.016
Aves										
Galliformes	*Gallus gallus*	Gryzinska et al. (2013)	Blood	1 day - 32 weeks	Global	-	Neg			Methylation level: 29.89% (chick) - 18.56% (adult), p = 0.036
		Sun et al. (2014)	Abdominal fat	2, 3 and 7 weeks	PPARγ	6	Neg			% methylation of 6 CpG sites in promoter associated with age (p < 0.0001)
		Zhang et al. (2017)	Breast tissue	20 and 55 weeks	Global (selected DMRs)	-	Pos			Increased global DNAm levels in 55-week-old versus 20-week-old across 2714 differentially methylated regions.
					ABCA1	Promoter	Pos			358 differentially methylated genes, 334 Pos, 45 Neg.

Order	Species	Reference	Tissue	Age range	Gene	CpG (N or ID)	Direction	MAD or error	MAD percentage of max. lifespan	Age relationship
					COL6A1	Promoter	Pos			
					GSTT1L	Promoter	Pos			
	Coturnix japonica	Andraszek et al. (2014)	Spermatocytes	15 & 52 weeks	RN28S	na	Neg			Higher intensity band in 15 wk old animals (MSP: qualitative)
Reptilia										
Crocodilia	*Alligator mississippiensis*	Nilsen et al. (2016)	Blood	Snout-vent-length: (sub-adult and adult)	Global	na	Neg			$R^2 = 0.04$, $p = 0.03$
Actinopterygii										
Cypriniformes	*Danio rerio*	Shimoda et al. (2014)	Muscle	2-day embryo, 3, 18 and 30 months	Global	na	Neg			Age-dependent genomic hypomethylation visualised on agarose gel following enzyme cleavage.
			Brain, liver	1, 3 & 30 months	EF1α2	na	na (liver), Neg (brain)			$p < 0.05$ (1m & 3m, liver), $p < 0.05$ (all ages, brain)
					FLI1A	na	Neg (liver), na (brain)			$p < 0.05$ (1m & 3m, liver)
					FGF3	na	Neg (both)			$p < 0.05$ (all ages, liver and brain)

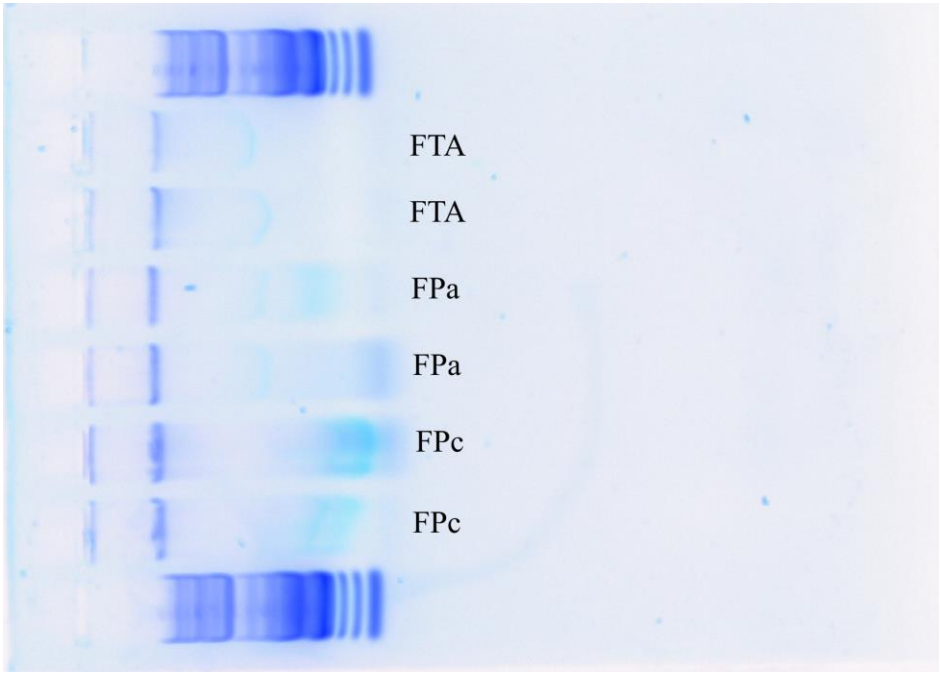
*Note that for humans and mouse studies the average lifespan is shown.

Supplementary Material

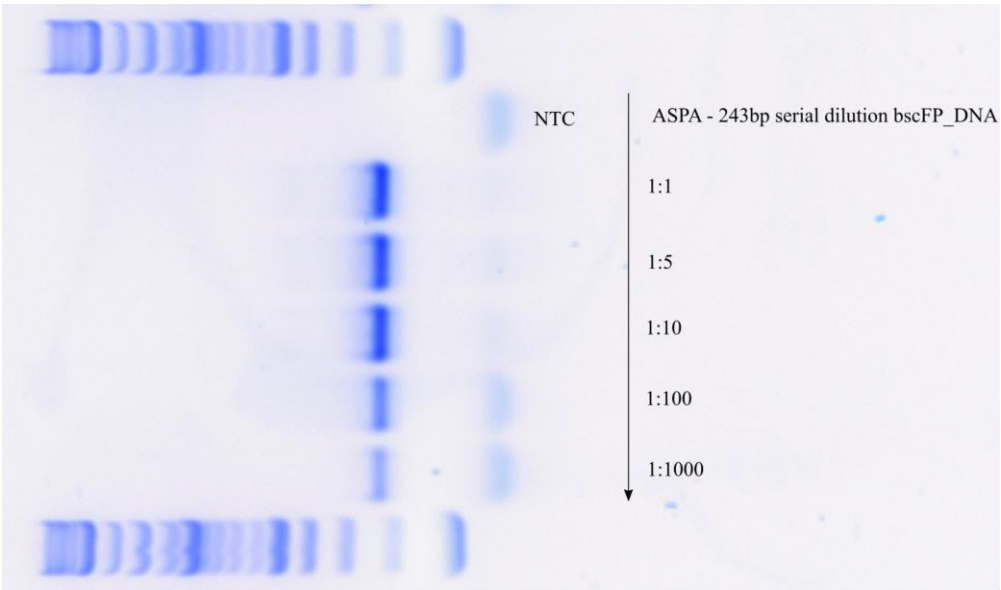
Chapter 3



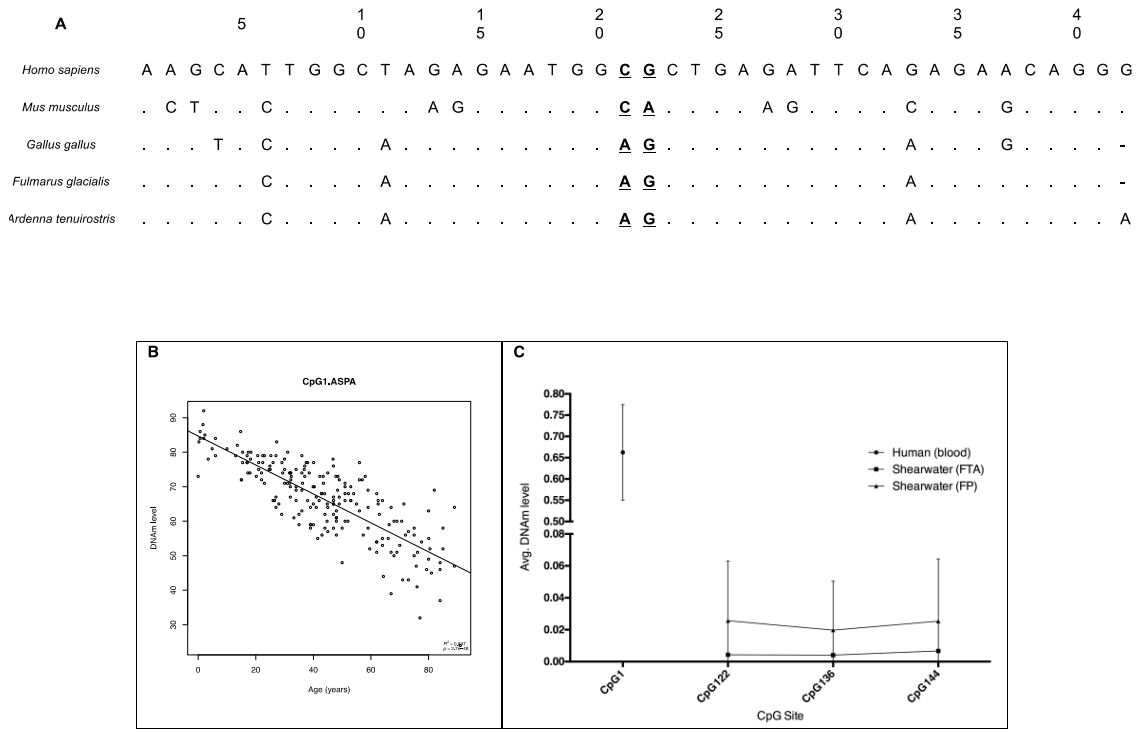
S.Figure 1 | qPCR sexing assay of *A. tenuirostris*. DNA blood stored on FTA cards was used to amplify a small amplicon in the CHD1 gene. Melt curve analysis reveals a double peak for females (n = 4, green) and a single peak for males (n = 7, red). A NTC (n = 1, blue) shows no contamination. A third primer dimer peak is also seen for female samples at approximately 72°C.



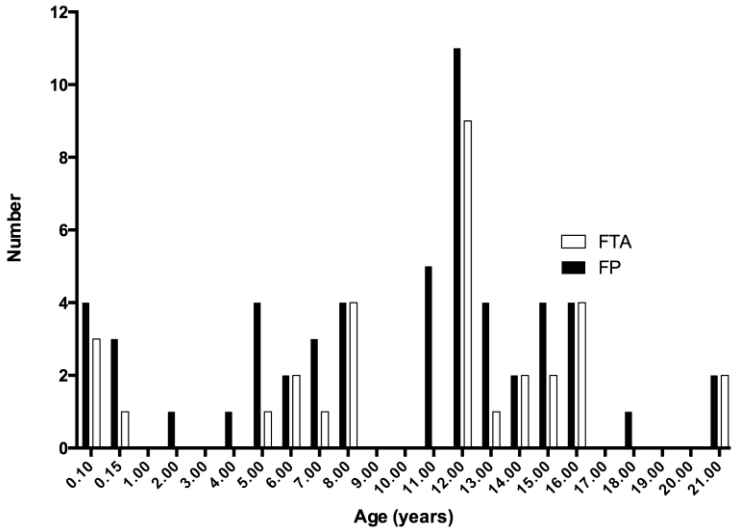
S.Figure 2 | Isolated DNA from blood and feather of *A. tenuirostris*. A QC gel indicating the isolation of high quality genomic DNA from shearwater tissue samples. FTA: DNA isolated from blood stored on FTA cards, FPa: plucked breast feather from adult, FPc: plucked breast feather from chick. Approximately 50 ng of DNA was loaded into a 1% agarose gel and run at 80 V for 30 minutes.



S.Figure 3 | Bisulphite converted PCR. A serial dilution of bisulphite converted DNA isolated from a plucked breast feather. In this example a 243 bp target region of the predicted ASPA gene in the shearwater was amplified using bisulphite specific primers. 100 ng of genomic DNA was bisulphite converted and eluted in 10 µL of elution buffer. At an approximate original concentration of 10 ng/µL this converted DNA was then serially diluted to 1:1000 and 1 µL was used in the reaction mix. A 2% Agarose gel running at 100 V for 50 minutes was used to visualise the resulting amplicons.



S.Figure 4 | ASPA sequence conservation in mammals and select bird species. A: 20 base pairs are shown for 3' and 5' directions around an age-related CpG loci (27K: cg02228185, in bold) in the ASPA gene in humans, mouse and birds. Following the methods described in text, this sequence was compared to the chicken and northern fulmar genomes. The resulting matches were aligned with the human sequence, conserved bases are shown with a dot (.), missing bases with a dash (-), and the base is given where there is a mismatch. In this example, the age-related CpG site in humans was not conserved in any other species. Primers were designed from the fulmar sequence to amplify isolated shearwater DNA to examine other CpG sites in the gene. **B:** The correlation of ASPA DNA methylation at this site [38] compared to **C:** the average methylation levels observed in shearwater blood and feather.



S.Figure 5 | Age distribution of *A. tenuirostris* used in this study. The known ages of all birds used in this study are shown split into those used for feather quill tip (FP) and whole blood (FTA) methylation analysis.

Supplementary File 1. Tables A – K.

Table A Northern fulmar designed genomic primers

Gene	Primer Num	Forward	Reverse	CpG_count	Size (bp)	450K_ID
Gene Set 1						
ASPA	1	CTTGGCCATGGTTTAGCTG	TTGTCCTGTCAAAGATCACCA	4	242	cg02228185
DNMT3A	1	TTCATGGAAGGACTGGGAAC	TTCTTGTTTGCTACCAGCTGTC	5	167	
EDARADD	1	AGGCCTCCTGCCATCCT	TCTGAGGTCCTCTGGTGTCC	4	170	
ELOVL2	1	TTCCATTTGATTTTATGAGTTTTCTG	TTTTTCCACTGATGCTGTGTG	7	250	
GRIA2	1	TGAGGATCCTTTAGGTGTGG	TGCAGAACCATCTGTGTTTGT	5	196	cg16777106
KCNC3	1	TGAACTTCATTGGATTGCTTTG	TGGCAAATCCATGTAGACACA	20	290	cg06572160
KCNQ1DN	1	GCTGGAGTGGTGGTAGCTGT	AAATGCTAGTCTGCAGCATCC	3	212	
NPTX2	1	TCAGAGGGTAGCCTTGTGGA	TTTTATCCCACCATTTTGCAT	13	175	
NPTX2	2	ATTCCTCCACTCCTCCTGCT	TGCTTTCACTGACATCTTGTTG	8	101	
PDE4C	1	TCCTGTCCCTTGTCACTGGT	GAGTGTCTGGAGGTGCTGGT	6	219	
TET2	1	GAAAGCCATGGTGCTGCAG	CATGGTGTCCTTGCCCTTG	5	227	
TET2	2	AAGAAGGAGCTGGGCAAGG	TCTTTCAGGGTCTGCATGGT	5	175	
TOM1	1	GAGCACCTGCACCACCA	AGGGTCTCAAAGGCCTCATT	5	178	
Gene Set 2						
EDARADD	2	CCCAAGATTATCAGGGTCAAA	GCATGACTTCCTCTTCTGCTG	33	285	
ELOVL2	2	GCCAGCACTCATCACTCAGA	GCTGGGCTGACTGCTATTTT	16	368	
ELOVL2	3	CTGCTTGATGGCTCACAATA	CCTCACCTGGGGGCA	29	231	
MYOD1	1	CACAGCCAAAGTGCTGATTC	GAGCAACAAGCATACCAGCA	13	209	cg18555440
TET2	3	CCCAGAGCAACTTTGACTTTGA	CTCTTCAGGCTCCTCTCCAC	9	327	
TET2	4	GAGTTTGACAAGTTTCTGGAAGA	TGCTCCACAGCATACCAGTC	7	237	
TRIM59	1	GGGAAGGATGTCTGTTGGAA	TGCCCTCTCTCTCACCTCAC	42	320	

Short-tailed shearwater (*Ardenna tenuirostris*) bisulphite converted primers

Table B

Gene	Primer Num	Forward	Reverse	CpG_count	Size (bp)	450K_ID
Gene Set 1						
ASPA	1	TATAGTTTTTGGGTTGGTAAGAAATG	AACTTCCTCTTCTACTATTAATAAACCTCC	3	175	
EDARADD	1	AAATTATGTATATAGTAGGTTTAAGGTATG	AAACTAACTACTATTTTCAATTCCTAC	4	116	
ELOVL2	1	TTTGGTTATGGTTTAGTTGAATGTGT	AACAAAAAAAACCAAAATCTTTCTC	3	225	
GRIA2	1	ATGGTATTATAGGGTTTTTTTTGTT	ACAAAACCATCTATATTTATAAAAACTACA	3	134	
KCNC3	1	TAATTTTGATTTTGATGGTAAAAGTAATG	ATAAATCATCCAACAACAAAACTCC	12	196	
PDE4C	1	TTTGATTTTATGAGTTTTTTGATTAATGTAATTG	TTAAAAATACAAACTTAAACAATACAAAAATACAACTC	7	270	
TET2	1	GTGTTTGTTGTTTTTGTTTTTGAAA	TAATCTACAACATCCCTACCCAAAC	3	185	
TET2	2	TGTTTTTTAGTTGTTTAGTTGATGGA	AACCCCTAAAAATTTTCAAAAACATA	6	185	
TOM1	1	GGAGAGTTAGAGGATGTAAAAGG	TTACTCCACAACATACCAATCTATAACATTCAC	7	180	
Gene Set 2						
ELOVL2	2.1	TTTATGGAAGGATTGGGAATTTAT	AACCTCCTTCCCATATAACAAAAATA	5	156	
ELOVL2	2.2	TGGGAAGGAGGTTATAATTTGTAGT	AAACCCTCATATTAAAAATATCATCAATAA	9	190	
MYOD1	1	AAGGTTTGTAAGAGGAAGATTATTAA	CTTAAAAATCTCAAAAACCTCATTAAC	7	105	cg18555440
PDE4C	2	TTGATTAATGTAATTGTAGGTTTTT	TTACCTATAATTTTTCCACTAATACTATAT	6	206	
TET2	3	GGAATAGGATAGAATTAATTATGTTGA	ATACTTTCTTTTCTTATAAAAAATCAAAAC	6	254	
TET2	4	AATGTTTTATAAAGATTTTTTTATTT	TTACTTTCACTAACATCTTATTATAAC	9	249	

Table C Genomic PCR cycling conditions

Temperature (°C)	Time	Cycles	Step
98	30 sec		1
98	10 sec		
56-67	30 sec	x40	2
72	30 sec		
72	5 min		3
4	∞		4

Table D Bisulphite converted PCR conditions

Temperature (°C)	Time	Cycles	Step
95	10 min		1
95	30 sec		
50-61	30 sec	x45	2
72	30 sec		
72	7 min		3
4	∞		4

Table E Sample barcoding cycling

Temperature (°C)	Time	Cycles	Step
95	10 min		1
95	30 sec		
55	30 sec	x10	2
72	30 sec		
72	7 min		3
4	∞		4

Table F DNA yield from analysed blood and feather quills

Tissue	FTA (adult)	Feather (adult)	Feather (chick)
N	30	47	6
Mean (ng)	1826.3	124.2	1844.0
SD	1295.1	116.7	343.5

Table G DNA yield from adult feather quills

Quills used	1	2	3
N	23	25	9
Mean (ng)	68.54	149.4	178.8
SD	34.47	141.3	70.4

Table H Absolute difference between feather replicates for gene set 1

Technical replicate	**ASPA**	**EDARADD**	**GRIA2**	**KCNC3**	**TET2**	**Total (%)**
Min	0.089729571	0.005342992	0.045689637	NA	0.006358778	
Max	0.10044166	0.051533101	0.078806476	NA	0.103619019	
Mean	0.094966216	0.024259353	0.067275047	NA	0.047471692	**5.849307688**
Run replicate						
Min	0.000336726	0.005616315	0.008073982	0.003833712	0.001064454	
Max	0.020742494	0.054545321	0.122303691	0.488451935	0.225571092	
Mean	0.007147183	0.033614236	0.051519817	0.115219413	0.044628794	**5.042588844**

Table I. Simple linear regression statistics for gene set 1.

Gene	CpG_ID	Tissue	R^2	Slope	*p*
KCNC3	66	FTA	0.325	+	**0.0190**
KCNC3	87	FTA	0.240	+	**0.0430**
KCNC3	31r	FTA	0.200	+	0.0600
KCNC3	82	FTA	0.190	+	0.0670
EDARADD	46	FP	0.168	+	**0.0067**
KCNC3	39	FTA	0.167	+	0.0820
ASPA	122	FP	0.157	-	**0.0081**
KCNC3	117	FTA	0.141	+	0.1000
ASPA	144	FP	0.140	-	**0.0120**
GRIA2	87	FTA	0.137	-	0.1100
GRIA2	81	FP	0.114	-	**0.0230**
KCNC3	129	FTA	0.111	+	0.1300
EDARADD	88	FP	0.088	+	**0.0420**
EDARADD	69	FP	0.076	+	0.0550
ASPA	122	FTA	0.074	+	0.1800
ASPA	144	FTA	0.070	-	0.1900
EDARADD	73	FP	0.068	+	0.0650
KCNC3	66	FP	0.068	+	0.0710
ASPA	136	FP	0.067	-	0.0640
KCNC3	39r	FTA	0.059	+	0.2000
TET2.2	74	FTA	0.051	+	0.2300
TET2.2	76	FTA	0.047	+	0.2300
KCNC3	107	FTA	0.033	+	0.2500
KCNC3	117	FP	0.009	+	0.2600
EDARADD	73	FTA	0.001	-	0.3400
KCNC3	87	FP	0.001	+	0.3100
KCNC3	43	FTA	-0.004	+	0.3500
GRIA2	81	FTA	-0.006	+	0.3600
GRIA2	25	FP	-0.014	-	0.4900
ELOVL2	37	FTA	-0.016	+	0.3900
EDARADD	69	FTA	-0.017	-	0.3900
KCNC3	27r	FTA	-0.017	+	0.3900
KCNC3	43	FP	-0.019	+	0.5500
KCNC3	82	FP	-0.019	+	0.5500
KCNC3	107	FP	-0.020	+	0.5600
KCNC3	129	FP	-0.020	-	0.5700
TET2.2	26	FP	-0.022	+	0.6300
TET2.2	40	FP	-0.023	+	0.6600
KCNC3	39	FP	-0.024	+	0.6600
TET2.2	96	FP	-0.024	+	0.7000
TET2.2	96	FTA	-0.024	+	0.4200

Gene	CpG_ID	Tissue	R^2	Slope	*p*
KCNC3	31r	FP	-0.024	+	0.6600
TET2.2	36	FP	-0.025	+	0.7100
GRIA2	87	FP	-0.026	-	0.7500
KCNC3	51	FP	-0.026	-	0.7100
TET2.1	61r	FP	-0.026	-	0.7400
KCNC3	27r	FP	-0.026	-	0.7200
TET2.2	76	FP	-0.027	+	0.8400
TET2.1	58r	FP	-0.028	-	0.8200
TET2.1	74r	FP	-0.028	-	0.8500
TET2.2	74	FP	-0.028	-	0.9200
KCNC3	39r	FP	-0.030	-	0.8800
TET2.1	58r	FTA	-0.030	-	0.4400
GRIA2	25	FTA	-0.034	+	0.4600
TET2.1	74r	FTA	-0.044	-	0.5100
TET2.2	40	FTA	-0.044	+	0.5000
KCNC3	51	FTA	-0.049	+	0.5400
ASPA	136	FTA	-0.050	+	0.5500
TET2.1	61r	FTA	-0.053	-	0.5700
EDARADD	46	FTA	-0.059	-	0.5500
TET2.2	26	FTA	-0.060	+	0.5800
ELOVL2	82	FTA	-0.070	+	0.7000
EDARADD	88	FTA	-0.071	-	0.6100
TET2.2	36	FTA	-0.076	+	0.7000
ELOVL2	71r	FTA	-0.079	+	0.8200

Table J. Simple linear regression statistics for gene set 2.

Gene	CpG_ID	Tissue	R^2	Slope	*p*
ELOVL2.2	42	FP	0.285	-	**0.0005**
ELOVL2.2	125	FTA	0.119	-	**0.0410**
ELOVL2.2	54	FP	0.115	-	**0.0240**
PDE4C	83	FTA	0.112	+	**0.0420**
MYOD1	67	FTA	0.111	-	0.0830
TET2.3	94r	FTA	0.102	-	**0.0450**
MYOD1	30	FTA	0.092	-	0.1000
ELOVL2.2	114	FP	0.070	-	0.0740
MYOD1	30	FP	0.067	+	0.0540
ELOVL2.2	122	FP	0.064	-	0.0840
ELOVL2.1	79	FP	0.050	-	0.0930
ELOVL2.2	54	FTA	0.045	+	0.1300
PDE4C	45	FP	0.039	+	0.1100
TET2.3	94r	FP	0.038	-	0.1400
TET2.3	57	FTA	0.038	-	0.1500
ELOVL2.1	89	FP	0.035	-	0.1400
PDE4C	38r	FTA	0.028	+	0.1900
TET2.3	37r	FTA	0.027	-	0.1900
PDE4C	93	FTA	0.024	+	0.2100
PDE4C	45	FTA	0.024	+	0.2100
MYOD1	61	FTA	0.018	+	0.2600
ELOVL2.2	144	FTA	0.017	-	0.2400
ELOVL2.1	89	FTA	0.011	-	0.2800
MYOD1	37	FP	0.006	-	0.2700
ELOVL2.1	25	FP	0.001	+	0.3200
TET2.3	35r	FP	0.000	-	0.3200
ELOVL2.2	116	FP	-0.007	-	0.3800
ELOVL2.2	125	FP	-0.009	-	0.2600
ELOVL2.2	114	FTA	-0.009	-	0.4000
ELOVL2.2	42	FTA	-0.011	-	0.4200
ELOVL2.2	127	FP	-0.011	+	0.4200
MYOD1	64	FP	-0.012	-	0.4800
TET2.3	70	FTA	-0.012	-	0.4300
MYOD1	34	FP	-0.013	+	0.4900
PDE4C	137	FTA	-0.013	-	0.4300
MYOD1	61	FP	-0.014	+	0.5100
PDE4C	26	FTA	-0.015	-	0.4500
ELOVL2.2	61	FTA	-0.015	-	0.4600
ELOVL2.1	79	FTA	-0.017	+	0.4800
TET2.3	28	FTA	-0.018	-	0.5000
PDE4C	93	FP	-0.019	-	0.5900
PDE4C	26	FP	-0.019	+	0.6100
MYOD1	27	FP	-0.022	-	0.7300

Gene	CpG_ID	Tissue	R^2	Slope	*p*
TET2.3	57	FP	-0.023	-	0.6300
MYOD1	67	FP	-0.024	-	0.8400
PDE4C	83	FP	-0.024	+	0.7600
ELOVL2.1	56	FP	-0.024	-	0.7500
PDE4C	38r	FP	-0.025	-	0.8100
ELOVL2.1	25	FTA	-0.025	-	0.6100
ELOVL2.1	33	FP	-0.025	+	0.9100
ELOVL2.2	61	FP	-0.026	-	0.1700
TET2.3	37r	FP	-0.026	+	0.7100
PDE4C	137	FP	-0.026	-	0.9800
ELOVL2.2	116	FTA	-0.027	-	0.5900
ELOVL2.1	56	FTA	-0.027	+	0.5800
TET2.3	35r	FTA	-0.028	+	0.6700
ELOVL2.1	33	FTA	-0.028	+	0.6800
TET2.3	28	FP	-0.030	+	0.9200
TET2.3	70	FP	-0.030	+	0.9800
MYOD1	37	FTA	-0.031	-	0.5200
ELOVL2.2	144	FP	-0.033	-	0.9900
ELOVL2.2	122	FTA	-0.036	+	0.7500
ELOVL2.2	127	FTA	-0.042	-	0.9500
MYOD1	27	FTA	-0.050	+	0.7600
MYOD1	64	FTA	-0.052	-	0.8100
MYOD1	34	FTA	-0.054	+	0.8600

Table K Human and fulmar gene conservation

Fulmar gene (predicted)	Human gene	Human age associated CpG ID	Query cover	E value	Identity	Chicken gene	Fulmar accession ID	Comments
ASPA	ASPA	cg02228185	99%	5.00E-29	83%	ASPA	XM_009579688.1	
ELOVL2	ELOVL2	cg16867657, cg16323298, cg01799681, cg24724428, cg21572722, cg21649660, cg05446010, cg22143569	53%	7.00E-130	77%	ELOVL2	XM_009579632.1	Some human conservation for exon base pairs 1000-2000 therefore chicken ELOVL2 sequence used for primers.
PDE4B	PDE4C	cg17861230	73%	2.00E-48	80%	PDE4B	XM_009581360.1	Human PDE4C exon region 2800-3500 has high similarity with chicken and fulmar regions predicted as PDE4B.
EDARADD	EDARADD	cg09809672	41%	4.00E-68	78%	EDARADD	XM_009578461.1	Poor human sequence conservation with fulmar, chicken sequence used to design primers for last exon.
GRIA2	GRIA2	cg25148589, cg16777106, cg17605476, cg22597733, cg00463631	69%	5.00E-87	89%	GRIA2	XM_009574686.1	
KCNC4	KCNC3	cg06572160	72%	1.00E-58	75%	KCNC4	XM_009582051.1	Human KCNC3 exon 1 matched highly with fulmar scaffold 21887 with a conserved CpG (predicted KCNC4)
TET2	TET2	cg02382073	73%	0	73%	TET2	XM_009575157.1	
MYOD1	MYOD1	cg18555440	57%	7.00E-65	86%	MYOD1	XM_009580318.1	This CpG ID associated with age in mouse.

Supplementary File 2. Python script used for calculating methylation proportion of specific CpG sites.

```
#! /usr/bin/python

# Script that reads in folder of fasta files and calculates % methylation @ given CpG (index) positions.

# DNA methylation results are exported to a .csv file. Read metrics are exported to a separate .csv file.

import sys

import os

import subprocess

import csv

import argparse

indeces = []


# Get args from command line

argsIn=sys.argv[1:]

for a in range(len(argsIn)):

	if argsIn[a]=='-in':

		#filename = argsIn[a+1]

		rootdir = argsIn[a+1]


	if argsIn[a]=='-dir':

		direction = argsIn[a+1]


	if argsIn[a]=='-indeces':

		indeces = (argsIn[a+1])

		indeces = indeces.split(',')


	if argsIn[a]=='-core':

		core = argsIn[a+1]

#Get list of files to process from input directory

filelist=os.listdir(rootdir)

print filelist

#print filename
```

```
print direction

indeces = map(int, indeces)
print indeces

# FASTA function - Reads a .fasta text file and returns lists of names and sequences
def readFasta(text):

        lines=text.split('>')[1:]
        names=[]
        seqs=[]

        for a in lines:
                line=a.split('\n')
                name='>'+line[0]+'\n'
                seq=''.join(line[1:])+'\n'
                names.append(name)
                seqs.append(seq)

        return names,seqs

# Methylation function F + R
def calc_met(fasta, genei, core):
#e.g. ASPA list of index locations fwd and rvr be different - need separate lists

        # Define vars
        genex = 0                                       # Position in given amplicon index list
        per_met = []                                    # Create empty list to receive % methylation
        CG_counts = []                                  # Create empty list to receive CG count
        TG_counts = []                                  # Create empty list to receive TG count
        fail_counts = []                                # Create empty list to receive failed count
        total_counts = []                       # Create empty list to receive total count
```

```
gene_count = len(genei)                                 # Length of given list of CpG sites
print 'The CpG index count is : ' + str(gene_count)


while gene_count != 0:


                cg_count = 0.0                          # Floating pt. for accurate div later
                tg_count = 0.0
                non_count = 0.0
                total_count = 0.0                       # Total read count per index
                per_meti = 0.0


                for seq in fasta:


                        seq = align5prime(seq,core)
# Call alignment function
                        #print seq
                # Error checking
                        c_count = 0.0
                        t_count = 0.0


                        if direction == 'F':
                                if seq[genei[genex]] == 'C' and seq[genei[genex]+1] == 'G':
                # Check given sequence position.
                                        c_count = c_count + 1
                                                # Count Cs at specific location
                                        #print c_count


                                elif seq[genei[genex]] == 'T' and seq[genei[genex]+1] == 'G':
                                        t_count = t_count + 1
                                        # Count Ts at specific location
                                        #print t_count
                                else:
                                        #print 'Not CpG pair'      # Error report
                                        #print seq                 # The error sequence
```

```
                        non_count = non_count + 1


                # Base counting
                cg_count = cg_count+c_count    # Contin. count of CG
                #print cg_count                # Check counts
                tg_count = tg_count+t_count    # Contin. count of TG
                #print tg_count


        elif direction == 'R':
                if seq[genei[genex]] == 'G' and seq[genei[genex]-1] == 'C':
# Check given sequence position.
                        c_count = c_count + 1
                                # Count Cs at specific location
                        #print c_count


                elif seq[genei[genex]] == 'A' and seq[genei[genex]-1] == 'C':
                        t_count = t_count + 1
                                # Count Ts at specific location
                        #print t_count


                else:
                        #print 'Not CpG pair'      # Error report
                        non_count = non_count + 1


                # Base counting
                cg_count = cg_count+c_count    # Contin. count of CG
                #print cg_count                # Check counts
                tg_count = tg_count+t_count    # Contin. count of TG
                #print tg_count
        else:
                sys.exit('You messed up, check args')


# Calculate total number of reads for given index
total_count = cg_count+tg_count
```

```
                print 'There are', cg_count ,'CG reads'
                print 'There are', tg_count ,'TG reads'
                print 'There are', total_count ,'total reads'
                print 'There are', non_count ,'non CpG reads'

                CG_counts.append(cg_count)
                TG_counts.append(tg_count)
                fail_counts.append(non_count)
                total_counts.append(total_count)

                # Calculate % methylation for given index and add to list of results
                if total_count != 0:
                        per_meti = cg_count/total_count          # Beta value
                        per_met.append(per_meti)                 # Append value to list
                        print per_met
                else:
                        per_met.append(0)
                        print 'No CpGs at location'

                genex = genex + 1                # Move forward in the given index list
                gene_count = gene_count - 1
                # Repeat % calculation for each CpG site until given index list is 0
                print gene_count
                print 'End of a loop \n'

        return per_met, CG_counts, TG_counts, fail_counts, total_counts

#Function to align 5' sequence end
def align5prime(seq,core):
        i = seq.find(core)

        #If the core sequence is not present
```

```
        if i == -1:
                seq = seq                       # returnNone causes error in run - return seq unchanged
        #If the core sequence is offset from 0, trim excess bases, or return as is if already right
        if i > 0:
                seqr = seq[i:]
                #print len(seqr)
                #print indeces[-1]
                if len(seqr) <= indeces[-1]:
                        seq = seq
                else:
                        seq = seqr
                return seq


# Export methylation function
def export_results(met, name, dire):
        with open('result.csv', 'a') as f:              # append to existing file
                writer = csv.writer(f)
                namel = []
                namel.append(name[:])           # Make name something writerow likes
                #print name

                writer.writerow(namel)          # Write sample name
                writer.writerow(dire)           # Write direction
                writer.writerow(met)            # Write result

        print 'Exported results!'


# Export metrics
def export_metrics(cg, tg, fa, tt, name):
        with open('metrics.csv', 'a') as f:                     # append to existing file
                writer = csv.writer(f)
                namel = []
```

```
                namel.append(name[:])                   # Make name something writerow likes
                #print name

                writer.writerow(namel)                  # Write sample name
                writer.writerow(cg)                             # Write cg count
                writer.writerow(tg)                             # Write tg count
                writer.writerow(fa)                             # Write failed counts
                writer.writerow(tt)                             # Write total seqs

                print 'Exported metrics!'

for a in filelist:
        filename = rootdir+a
        print filename

        #Change the path to open the file you want
        fastaFile = open(filename)
# e.g. /Users/depaolir/Desktop/outtest/STS16001-ASPA_S92_L001_R1_001.fasta
        text = fastaFile.read()
        names,seqs = readFasta(text)
        print seqs[0:3]                                 # Test function of list
        print len(seqs)                                 # Total number of reads


        export = []                     # Empty lists to take return from methylation function
        cg = []
        tg = []
        fail = []
        total = []

        # Run calc_met function - assign return value to result list
        export, cg, tg, fail, total = calc_met(seqs, indeces, core)  # Change second arg to current
CpGs in target gene
```

```
#print export
# Results manipulation
res_list = len(export)
print 'There are ' + str(res_list) + ' methylation results'
results = []
results = list(indeces)                                    # Create copy of indeces for results list
i = 1
a = 0
while res_list != 0:
        results.insert(i,export[a])
        i = i+2
        a = a+1
        res_list = res_list - 1
        #print 'End loop \n'

print results
#print indeces
# Run export and metrics write to .csv
export_results(results, filename, direction)
export_metrics(cg,tg,fail,total,filename)
```

Supplementary File 3. R script that does linear regression for each CpG site and uses the package 'glmnet' for lasso predictor selection.

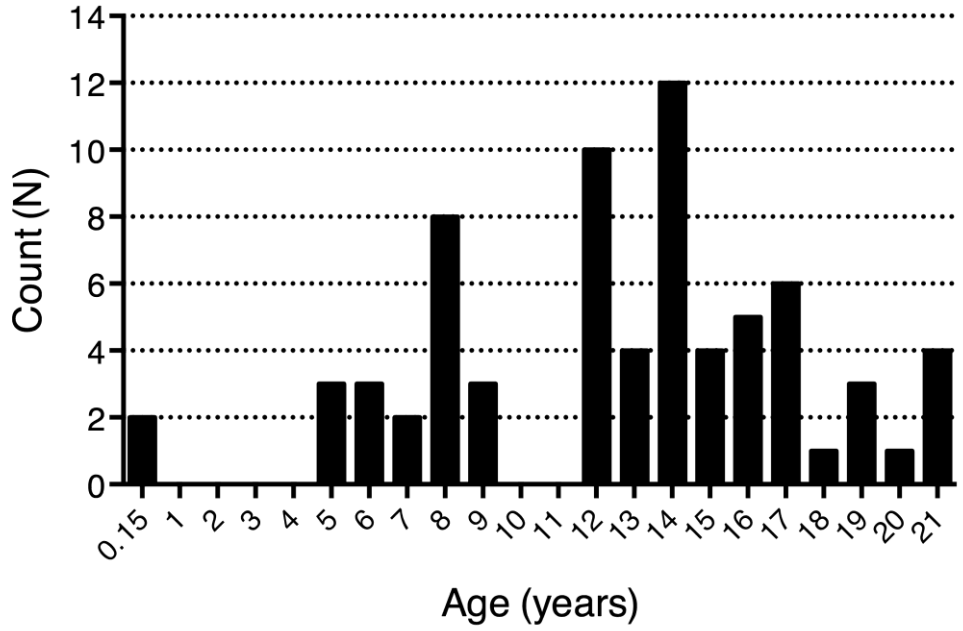
```
# Shearwater Epigenetic Age Assay (SEAA).
# Ricardo De Paoli-Iseppi - University of Tasmania, Australian Antarctic Division.
# This script takes in a .csv file with seabird age, sex and CpG loci methylation results between 0-1.

library(glmnet)
library(Matrix)
library(ggplot2)
#?glmnet

tdata=read.csv('/data.csv', na.strings=c("","  ","NA"), header = T)
head(tdata)
cc <- complete.cases(tdata)
age=tdata$Age
id=tdata$UID

# Plot age and ID
plot(age, id, main = "Individual ages", xlim = c(0,30), cex = 0.5)
text(age, id, labels = id, cex = 0.7, pos=4)

# Simple plots of age vs CpG loci.
pdf("grid_fig_test.pdf") # Save plots to a single .pdf.
nlocmax = ncol(tdata) + 1
loc = 3
colname <- colnames(tdata)
par(mfrow=c(3,3))

# <<Move .pdf line to here if not wanting grid.
while (loc != nlocmax) {
  lr1 <- lm(tdata[,loc] ~ age)
  modsum = summary(lr1)
  modsum$coefficients
```

```
  #plot(age,tdata[,loc], type = "p", cex = .6, xlab = "Age (years)", ylab = "DNAm level",
      #col = c("black"), main = colname[loc])
  plot(age,tdata[,loc], type = "p", cex = .6, xlab = "Age (years)", ylab = "DNAm level",
      col = c("blue","red", "green")[tdata$rep], main = colname[loc])
  abline(lr1)
  r2 = modsum$adj.r.squared
  my.p = modsum$coefficients[2,4]
  max = max(tdata[,loc], na.rm = TRUE)
  min = min(tdata[,loc], na.rm = TRUE)
  ymag = (max-min)
  #slope = modsum$coefficients[2,1]

  rp = vector('expression',3)
  rp[1] = substitute(expression(italic(R)^2 == MYVALUE),
                  list(MYVALUE = format(r2,dig=3)))[2]
  rp[2] = substitute(expression(italic(p) == MYOTHERVALUE),
                  list(MYOTHERVALUE = format(my.p, digits = 2)))[2]
  rp[3] = substitute(expression(italic(range) == MYOTHERVALUE2),
                  list(MYOTHERVALUE2 = format(ymag,dig=2)))[2]
  #rp[4] = substitute(expression(italic(m) == MYOTHERVALUE3),
                  #list(MYOTHERVALUE3 = format(slope,dig=2)))[2]


  legend('bottomright', cex=0.6, legend = rp, bty = 'n')
  loc = loc + 1
}

dev.off()

par(mfrow=c(1,1))
# Set up for glmnet pkg - lasso
X = as.matrix(tdata[cc,3:36]) # Change tdata to include only CpG data for specific row (all).
```

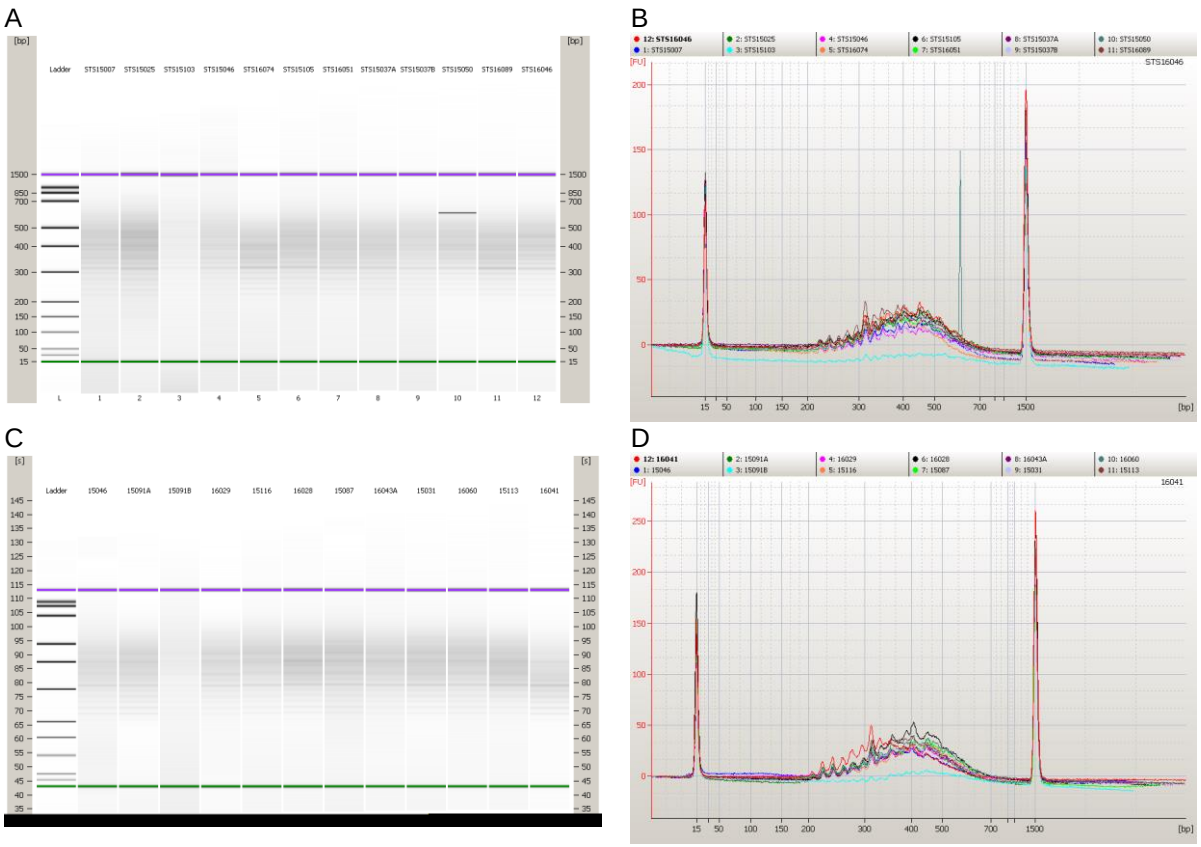
```
dim(X)
Y = tdata[cc,2] # All rows from column 2 (age)

# Lasso model using glmnet
fit=glmnet(X,Y, alpha = 1) # Fit age to CpG results
plot(fit) #L1 norm fit
plot(fit,xvar="lambda", main="SEAA model coefficient paths")
fit # Look at fit of each co-eff.

cv = cv.glmnet(X,Y, alpha=1)      # Perform cross validation on the fited model
plot(cv)
min <- cv$lambda.1se
coef(cv, s = "lambda.1se") # Or use '1se'

model = glmnet(X,Y,lambda=fit$lambda.1se)
plot(model)
plot(model,xvar="lambda", main="SEAA model coefficient paths")
predict(model, type="coefficients")
```

Supplementary Material

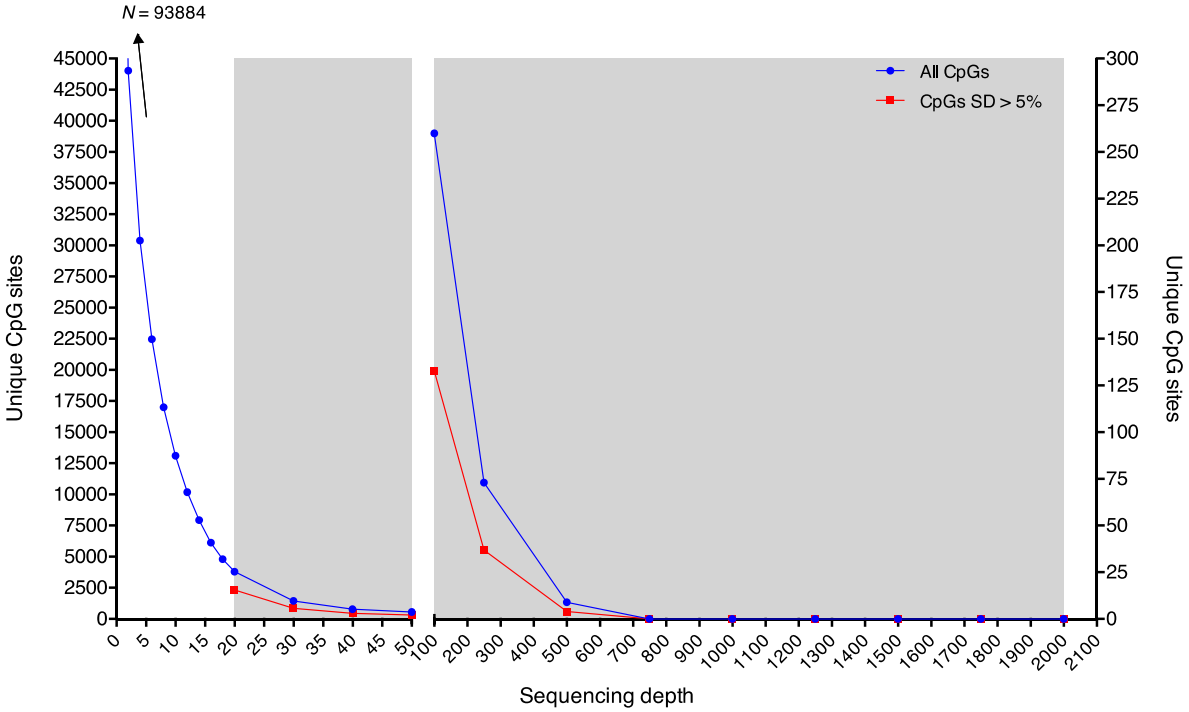
Chapter 4



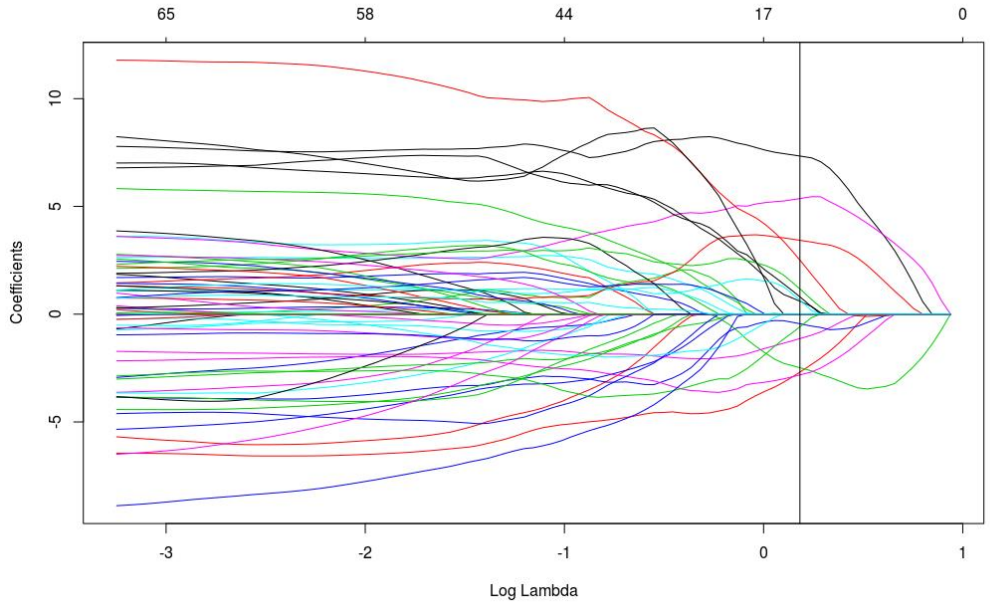
Supplementary Figure 1. Known-age distribution of all short-tailed shearwater blood samples used for DREAM analysis and model calibration and testing.



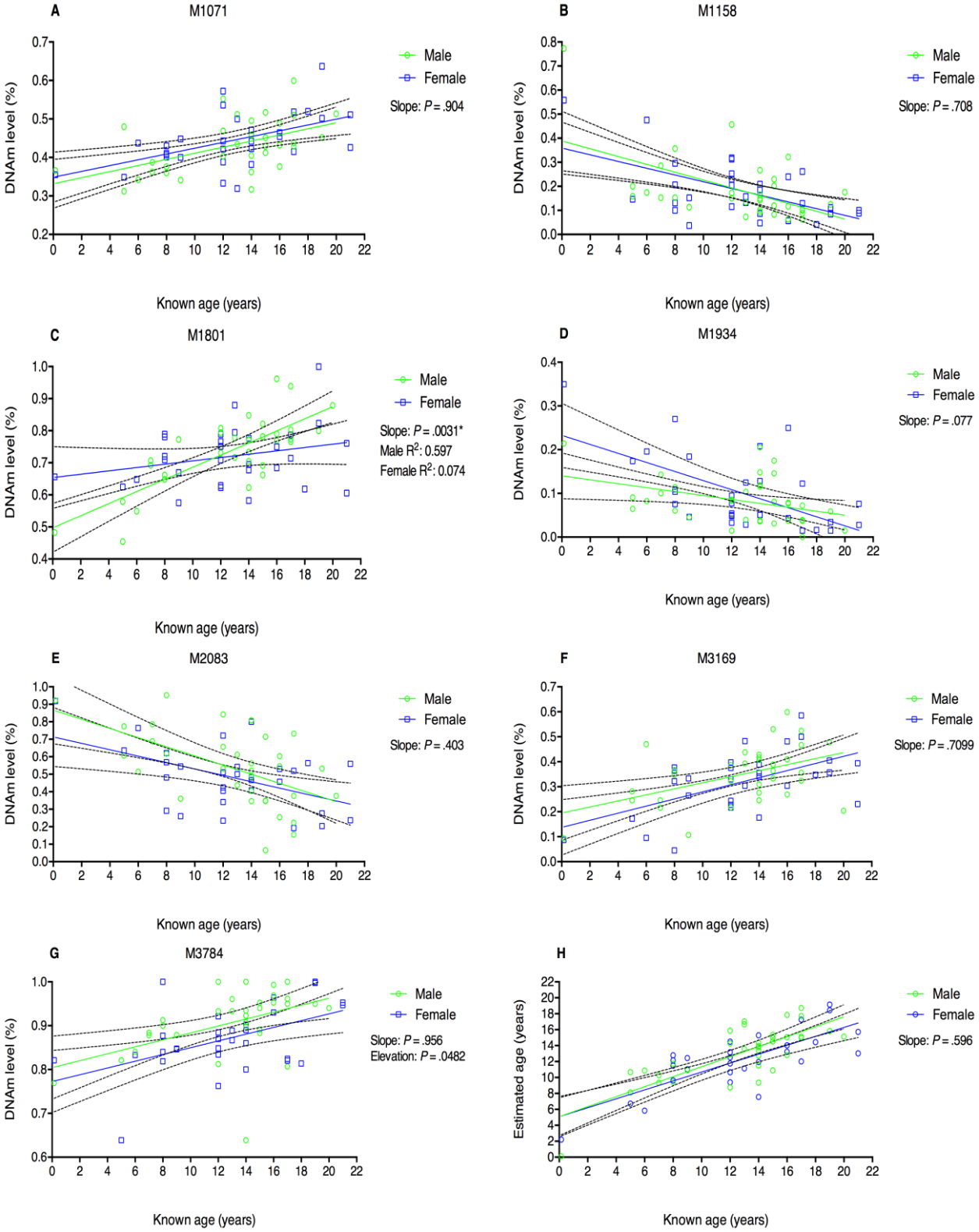
Supplementary Figure 2. Quality control of samples prior to DREAM library pooling. A. Example Bioanalyser artificial gel electrophoresis of samples from run 1. **B.** Example electropherogram trace of samples in run 1. **C.** Example gel electrophoresis of samples in run 2. **D.** Example electropherogram trace of samples in run 2.



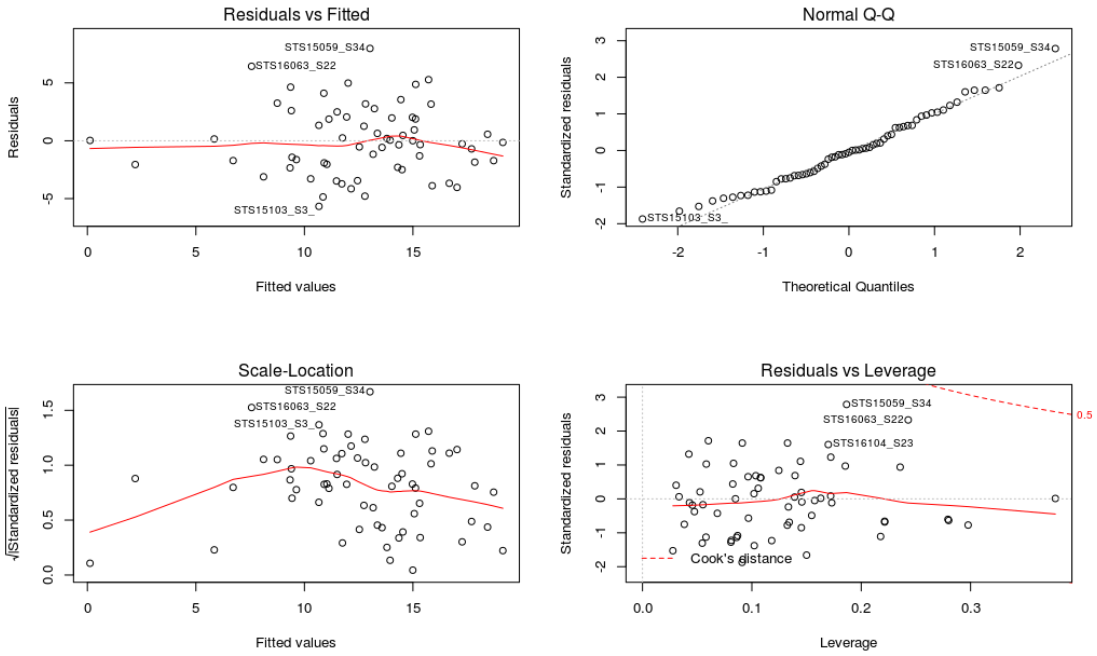
Supplementary Figure 3. Sequencing depth and unique CpG sites. Sequencing depth as done on the Illumina NextSeq platform for DREAM. The number of unique CpG sites identified is shown on the left and right *y-axes* for a diminishing window between the read depth indicated and 2000x reads per individual. If a CpG had > 7 read depth values outside the window it was dropped. The blue line indicates number of CpGs sites identified before filtering on SD of > 5% (shown in red from 20x read depth onwards). The shaded area represents the read depth range (20-2000x) at which DNAm data on CpGs was included for age analysis (*N* = 2338).



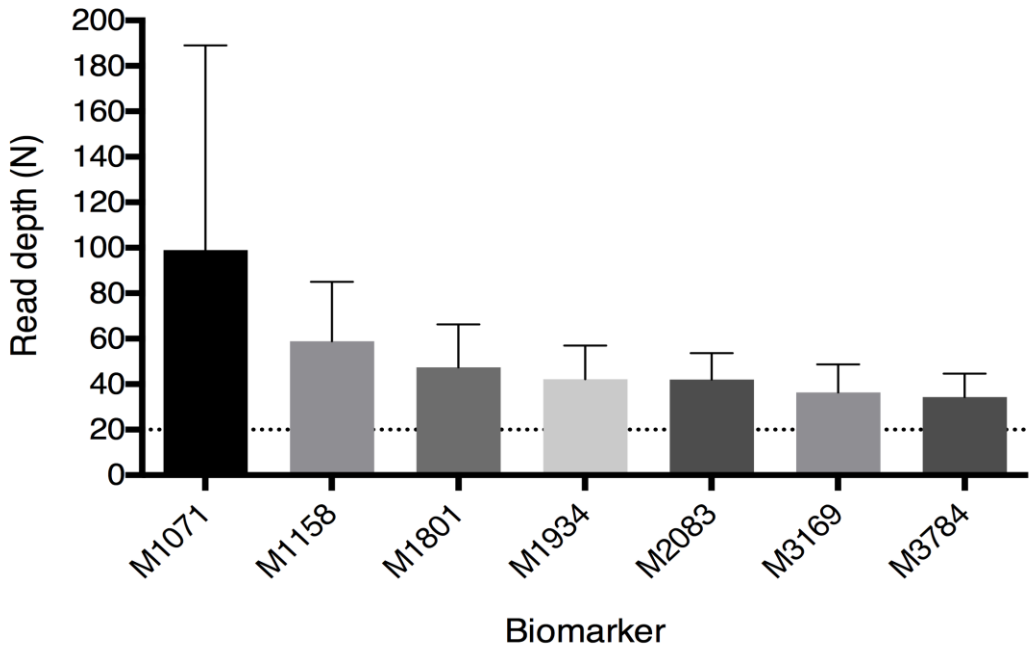
Supplementary Figure 4. Example coefficient pathways and lambda value cut-off. Coefficient pathways of a subset of CpGs identified using DREAM were selected based on sites that passed the lambda 1 standard error (**λ**1se) = 1.2 cut-off before penalisation out of the model using lasso (log **λ**1se is shown as a solid vertical line).



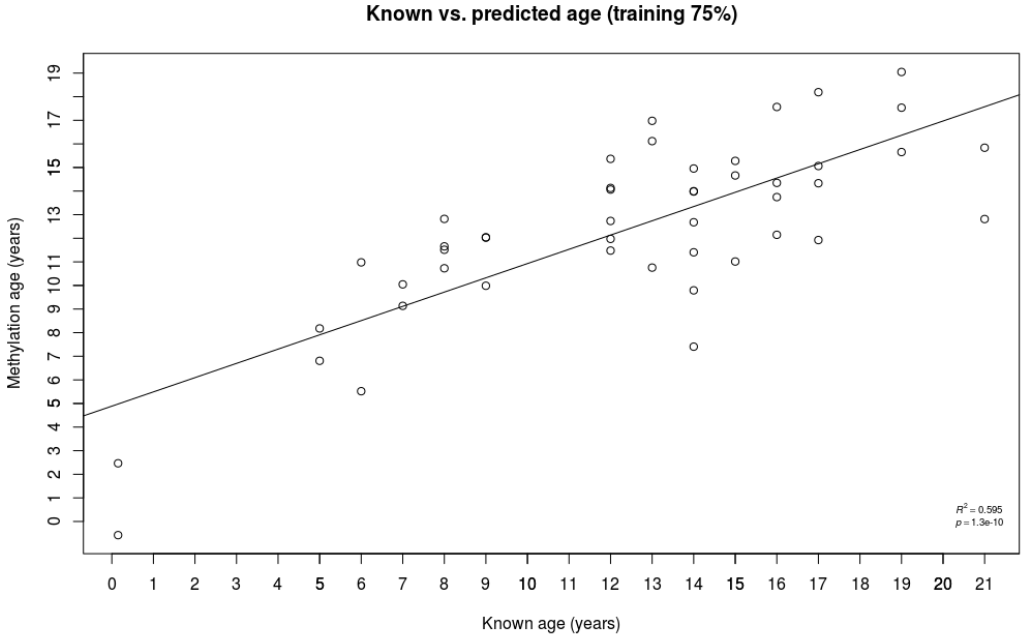
Supplementary Figure 5A – G. Age-related CpG sites showing known sex regression slopes. Linear regression of DNA methylation and chronological age for each sex and CpG selected using lasso penalisation from a total of *N* = 63 Short-tailed shearwater blood samples. Estimated age using DNAm scores for these markers are shown for each sex in **5H**. Males and females are shown in green and blue respectively for all figures in this set.

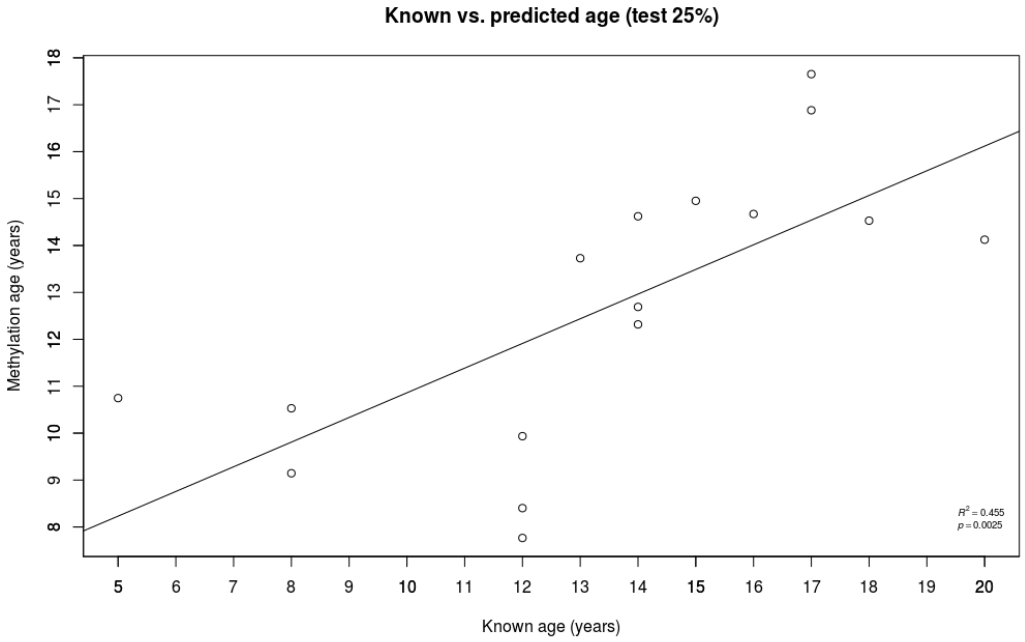


Supplementary Figure 6. Multiple linear regression diagnostic plots for full model. This set of figures identifies any potential problems with the CpG sites selected for the model, including individual outliers (Cook's distance) and normalcy (Q-Q plot).

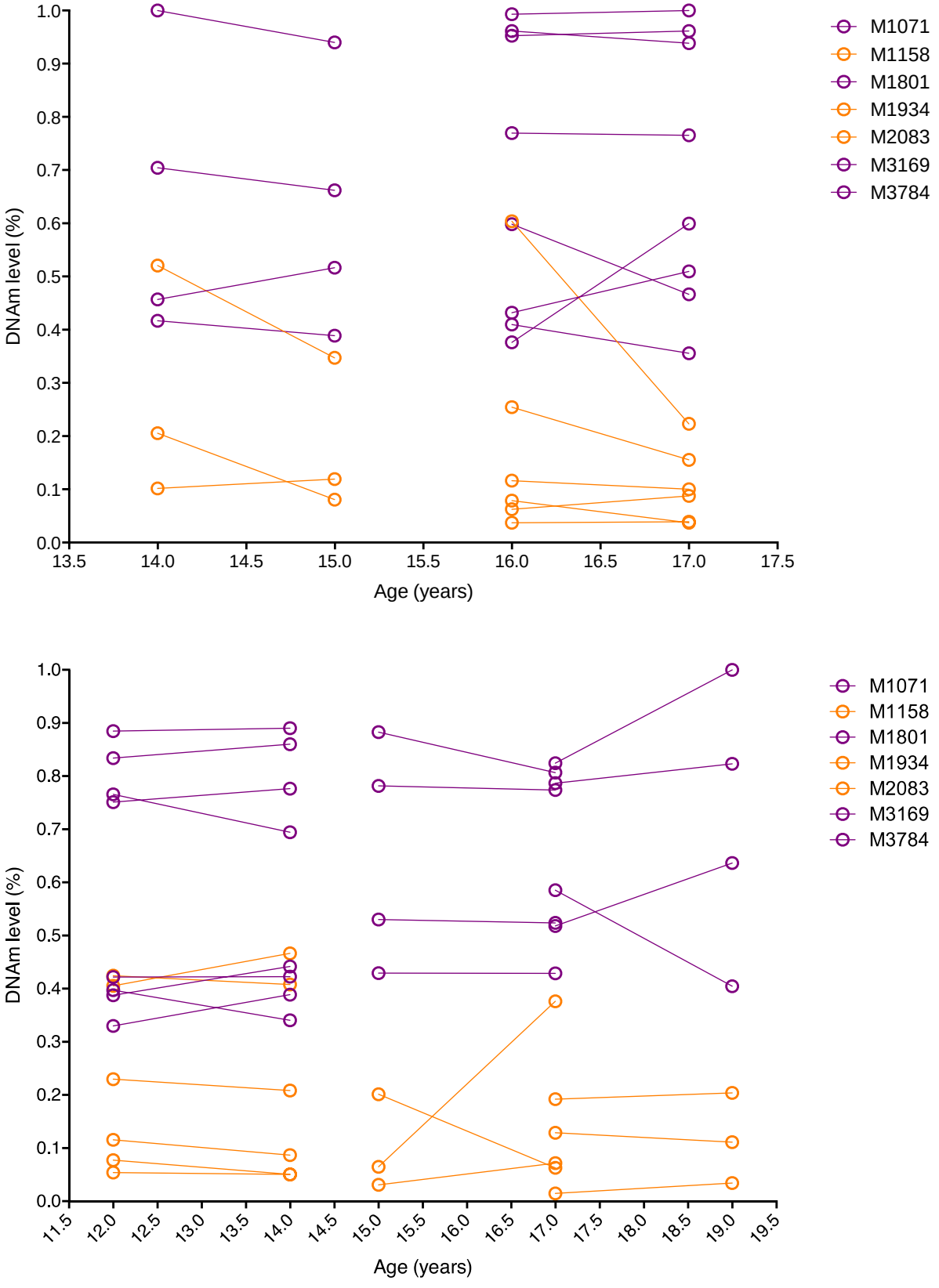


Supplementary Figure 7. Biomarker read depth statistics. The mean read depth used to calculate the DNA methylation score, ranging from unmethylated (0) to methylated (1), for each of the markers used in the age estimation model.





Supplementary Figure 8. Example of random training and test subset of predicted ages. An example of the seven aDMP predicted age (*y* axis) plotted against known-age (years) on the *x*-axis of **A.** training set (*N* = 47, 75%) and **B.** test set animals (*N* = 16, 25%). Linear regression shows correlations of $R^2 = .595$ and $R^2 = .455$ for these specific training and test subsets, respectively.



Supplementary Figure 9. Longitudinal DNA methylation for 1 and 2 year resights. DNAm scores for each of the resighted individuals (*N* = 7, 1-year *N* = 3, 2-year *N* = 4) are shown for each aDMP selected for the age model. aDMPs that show an overall increase or decrease in DNAm are shown in purple and orange, respectively. The 1-year resights are shown in **(A)**. The 2-year resights are shown in **(B)**.

Supplementary Table 1. Ligation-mediated PCR amplification (cycling conditions).

Cycles (x)	Temperature (°C)	Time
1	95	10 m
	55	30 s
12	55	30 s
	72	30 s
1	72	1 m
-	4	∞

Supplementary Table 2. Sequence QC and pipeline filtering.

Code order	Sample UID	Age	Run	Total seqs	Seqs with RE	Match CC%	Match GG%
1	STS15007	5	1	1926517	1425322	66.97	33.03
2	STS15013	16	1	2338253	1659858	72.65	27.35
3	STS15020	15	1	2324983	1772794	67.31	32.69
4	STS15025	5	1	2482894	1864107	69.98	30.02
5	STS15037	8	1	4446297	1466310	70.21	29.79
6	STS15037	8	1	2578971	1685816	72.57	27.43
7	STS15044	12	1	2130752	1640614	70.47	29.53
8	STS15046	9	1	2624869	1729839	68.4	31.6
9	STS15047	16	1	2553460	1805170	68.77	31.23
10	STS15050	8	1	2374789	1680075	69.18	30.82
11	STS15057	12	1	2418508	1679516	66.49	33.51
12	STS15059	21	1	2474273	1714323	66.55	33.45
13	STS15103	5	1	1613361	963165	68.11	31.89
14	STS15105	7	1	2603950	1503929	70.12	29.88
15	STS15111	12	1	2302868	1620282	69.85	30.15
16	STS16011	0.15	1	2438467	1708910	69.06	30.94

Code order	Sample UID	Age	Run	Total seqs	Seqs with RE	Match CC%	Match GG%
17	STS16020	0.15	1	2115935	1474770	70.44	29.56
18	STS16021	14	1	2048334	1365325	70.61	29.39
19	STS16021	14	1	2169207	1432822	70.3	29.7
20	STS16036	15	1	2347717	1859254	72.54	27.46
21	STS16044	17	1	2221700	1594204	69.36	30.64
22	STS16046	9	1	1740158	1320126	67.75	32.25
23	STS16051	7	1	2274856	1537312	67.94	32.06
24	STS16052	19	1	2301730	1715864	71.96	28.04
25	STS16058	13	1	2060299	1591537	67.47	32.53
26	STS16063	14	1	2494782	1562813	70.43	29.57
27	STS16067	18	1	2513999	1856476	65.14	34.86
28	STS16074	6	1	2251468	1714142	67.25	32.75
29	STS16075	13	1	2106110	1610492	69.54	30.46
30	STS16078	16	1	3182246	1581905	68.02	31.98
31	STS16080	9	1	1987694	1489218	65.67	34.33
32	STS16081	13	1	2977166	1514098	66.85	33.15
33	STS16084	17	1	2314047	1717847	67.22	32.78
34	STS16089	8	1	2014538	1503704	66.57	33.43
35	STS16104	14	1	2380965	1800188	65.71	34.29
1	STS15046	6	2	2564878	2046618	60.23	39.77
2	STS15091	8	2	2421349	1994176	52.96	47.04
3	STS15091	8	2	1261252	1037662	50.92	49.08
4	STS16029	8	2	2498223	2026476	56.16	43.84
5	STS17060	8	2	4980645	1678460	54.61	45.39
6	STS17001	9	2	5009542	1476920	55.63	44.37
7	STS15057	12	2	2591459	2075007	57.04	42.96
8	STS15065	12	2	2833266	2107045	54.35	45.65
9	STS15073	12	2	2730561	2144257	54.74	45.26

Code order	Sample UID	Age	Run	Total seqs	Seqs with RE	Match CC%	Match GG%
10	STS15075	12	2	2749753	2086569	56.78	43.22
11	STS15077	12	2	2441178	1898889	55.38	44.62
12	STS15099	12	2	2798557	2129951	57.94	42.06
13	STS15116	12	2	2897866	1893946	58.38	41.62
14	STS16028	12	2	2678544	2042944	59.71	40.29
15	STS15087	14	2	2851201	2189688	58.48	41.52
16	STS16043	14	2	2571260	2020822	58.64	41.36
17	STS16043	14	2	2238946	1873976	59.93	40.07
18	STS16102	14	2	2686660	1965757	59.45	40.55
19	STS17014	14	2	4689463	1642302	57.59	42.41
20	STS17015	14	2	5363036	1642824	57	43
21	STS17044	14	2	4282992	1941357	55.16	44.84
22	STS17046	14	2	3412212	1611531	56.64	43.36
23	STS15026	15	2	2872108	2172992	56.66	43.34
24	STS17030	15	2	4332706	2126045	54.27	45.73
25	STS15031	16	2	3585767	1744109	58.21	41.79
26	STS16060	16	2	2771839	1871607	58.85	41.15
27	STS15113	17	2	4381064	1751922	58.73	41.27
28	STS16041	17	2	4208514	2088041	53.67	46.33
29	STS17034	17	2	2997709	2251634	59.32	40.68
30	STS17052	17	2	2592878	1972692	56.16	43.84
31	STS15117	19	2	2510298	2001644	59.71	40.29
32	STS17059	19	2	2578512	2127656	62.2	37.8
33	STS17065	20	2	3286216	1775381	56.06	43.94
34	STS15059	21	2	2730334	2225545	54.08	45.92
35	STS17038	21	2	4055530	1853942	56.84	43.16
36	STS17038	21	2	5255260	1422650	55.85	44.15

Supplementary Table 3. Additional age-related CpG site sequences not included in multiple linear regression and their top BLASTn result (all species).

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
500	CCCGGGAACCAACAGGGCCGCCGGGCAGC CCCGTGCCCCAGGACTACAGCTCCCAGCAT GCCCTGGTGCACACCCTC	1se	0.189	Negative	*Nippon nippon (Crested ibis)*	MHC class 1 (genomic)	85	86	2.00E-09
1337	CCCGGGGCTTTGGGGCTGGGAACGACGCCT TCCGACCCCAGGACGCGACAGGGACCCCG CCGGGTGGGAGGAGGAAG	1se	0.225	Negative	*Low/poor conservation*	[redacted]			
1876	CCCGGGGATTATTTTTTTTTGTTTTAAAGTTA GTAATTGAAAACAATGTTAAAACTAGTCAGC GGGTGCGGGCTGAA	1se	0.209	Positive	*Calidris pugnax (Ruff)*	Uncharacterised ncRNA	90	94	9.00E-21
3144	CCCGGGGGTGGGCTCCGTCCCTGGCTGGG AGCGGGGCCGGGAGCGGGGCTGGGAGCAG GAAGAGCCGGGCCTGGCAG	1se	0.236	Negative	*Callorhinchus milii (Australian ghostshark)*	TNIK (mRNA)	80	83	6.00E-04
4295	CCCGGGCTGCCGGGCAGCACCTGCGGCTG CCGAAGGGAGACGGCCTTTGGTGGCGGGG AAGGGGCAGGAGGGCTGCA	1se	0.178	Positive	*Low/poor conservation*	[redacted]			
4336	CCCGGGCTGCCGCAGAAGCACCTAGGGCA GACTGGTGATGAGCAGAATGGAGCAGAGGG GCGGGGAAGAGCGGGCTC	1se	0.181	Negative	*Low/poor conservation*	[redacted]			
5788	CCCGGGCAGGTCCCATGGGTGATGCCGGC CCAAAATGGGGGCTCCCAGGTGCAGGGCTG TGCTCCGACCCTGTTTAG	1se	0.150	Positive	*Parus major (Great tit)*	FAM20C-like, transcript variant X1	59	91	1.00E-07
10982	CCCGGGGAAGAAAACGAGGGCCTGAGGAA GAAAACCAGAGCTCAAGGAACAAAACCTGA GCCCGAGGAACAAAAGAA	1se	0.185	Positive	*Nestor notabilis (Kea)*	DNAI2 (mRNA)	81	80	0.008
596	CCCGGGACATCTCCAGGGAGCCCAGAAGGG GCAGAGAAAGTGACCCCTTGGGCTGGTCCT CTGCTCCTGGGCTGGGC	min	0.166	Positive	*Corvus brachyrhynchos (American crow)*	Olfactory receptor 14I1-like (mRNA)	100	77	1.00E-05

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
784	CCCGGGGCGGCGGCGAGCGCGCGCGGCC GTCGCTGTCCCAGGACCGTGGCCGTGGGG CCCGCGCTTCCTTTCCGTCT	min	0.082	Positive	*Low/poor conservation*				
929	CCCGGGAGCGGGAGCGGGACTTGGAGAGG GAGCCGCTGGAGCTCCAGGAGCTGCTGCGC TGGCCGGGGCTGGGTGAC	min	0.069	Negative	*Egretta garzetta (Little egret)*	SRRM3 (mRNA)	100	99	2.00E-28
1127	CCCGGGAGCGGGCAGAAAGGGCCCTTTGA GAGGACACTCTGCTGGAATAAAAGTCACTTT TATCCCAGAATAATTAG	min	0.188	Negative	*Apteryx australis mantelli (North Island brown kiwi)*	scaffold1373	97	95	2.00E-23
1180	CCCGGGATTTTGTCCCCTGGGGGGGCTCCC TCCCAGGTCCCCGAGGGAGGGGTGCAGGG ACCAAGCGGTCGCTTCGT	min	0.083	Negative	*Low/poor conservation*				
1261	CCCGGGGAACAAAACCAGAGACCAAGGAAC AAAACCAGGGACTGAGGAACAAAACCAGAG CCCCAGGAAAAAAACCA	min	0.176	Positive	*Low/poor conservation*				
1374	CCCGGGGAACAAAACCAGAGCCCACAGGAG AAAACCAGAGCCCGCGGGAGAAAACCAGAG CCCGCGGGAACAAAACC	min	0.015	Positive	*Low/poor conservation*				
1482	CCCGGGTAGCAGAAGGTGACCGGAGCTCTG CAAACCTATTCGGAGGCCAAAAACGACCTTC TCTGTCCTACCTCTTG	min	0.033	Positive	*Low/poor conservation*				
1551	CCCGGGGACGGCCCCGCTTCGGCCCTGGC ACTGGGCGCATCTGGCTGGACGACGTCCGC TGCCGGGGTGAGGAGGGG	min	0.077	Positive	*Nippon nippon (Crested ibis)*	SSC5D (mRNA)	100	95	1.00E-24
1757	CCCGGGAGAGGGGCCACCACGGGGGACCC TCGCTGGGGCAGGCATGACAGCAGCAGGC GGAGAACGCGGGTGACGGC	min	0.043	Positive	*Low/poor conservation*				
1789	CCCGGGTTCCCGCCATATTAAAACAGAAGTT GGGGCGACAGCCGTCCCGGCGCGCGATCC CCCCATCCGTGCCGAGC	min	0.231	Positive	*Low/poor conservation*				
1847	CCCGGGGAAAGTGCGTGGGTCGGCGGATT GTTTTCTCTCTGCTGGAGCCTGGCGGGGAG GCCGGGGCCGGGGGCCGC	min	0.045	Positive	*Low/poor conservation*				
1924	CCCGGGAGTCTGGAAAGCTTCCTCCCCCCG GATGCGGGGCACGAGGGCCACCTTCGTGGA	min	0.147	Positive	*Aptenodytes forsteri (Emperor*	CNTNAP2 (mRNA)	100	92	8.00E-22

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
	GCACGAGCAGCAGGCAC				*penguin)*				
1939	CCCGGGAGGAGCCGGGGGAGCCGCAGCAT CGCTGCCTGCTGATGGACGTCCCCAGGGGC CCCTCGCTCCTGGGGCTG	min	0.080	Positive	*Low/poor conservation*	[redacted]			
1971	CCCGGGAGAGCCCGGTCTGTGGGGCGGGA GCTGTCGGGCCGGGGGCGAGGGTCCCCGG CCTCTCCTCTGAGAAGGGG	min	0.048	Negative	*Low/poor conservation*				
2078	CCCGGGCTCGCCCGCCGCTTGAAGTTGGAG GGAAAAGCGAGATCTAAGCAGCCCTTCTAAA AATGCATTAACTGTTT	min	0.018	Positive	*Apteryx australis mantelli (North Island brown kiwi)*	scaffold36	100	90	4.00E-19
2162	CCCGGGGGCACCCAGCTGCCGTACACCTCT GCCAGGCATTGACAGCCAGGGAGGGGGAT GTGCCCGCTCCGGGGCAC	min	0.087	Negative	*Low/poor conservation*	[redacted]			
2217	CCCGGGATGCTGCAGGAAACAAAGGGCTCT GCAAGCTCTGCCTTGTTGTTTCTAGCTGTGA AATTTACTTTTCTTTT	min	0.063	Negative	*Apteryx australis mantelli (North Island brown kiwi)*	scaffold1382	100	92	8.00E-22
2381	CCCGGGCAGGAAGGGAGGGAGCCGGGCAG GGCAGGACACGGCGCTCCGGCCGGCAAAG GAAGCGGCGTTTCCTGCGG	min	0.130	Positive	*Coturnix japonica (Japanese quail)*	BEST4, transcript variant X1 (mRNA)	94	89	2.00E-16
2687	CCCGGGAGGGCGGAGGGTGGATGTGGGGA GGCAGCTGAGCCTCATTGCAACCAGCTCGT CCGCTCCAGCAGCTCCGG	min	0.041	Positive	*Dromaius novaehollandiae (Emu)*	LPIN3, transcript variant X2 (mRNA)	58	84	0.008
2856	CCCGGGAGGAGGTGCGCCGCAGCCGCCGC TTGGCCTTGGCCTTCAGCCGCGCCTCGTGC AGCAGCTTCTCCTGGGGG	min	0.058	Negative	*Calidris pugnax (Ruff)*	Programmed cell death protein 4-like (mRNA)	100	99	2.00E-28
3006	CCCGGGAGGATAGATGGCACAGCCTTGCTA CAGCTCTCATGGCTCTTGTCCCACCCGCAC GCGGCGTAGCTTGGCTG	min	0.052	Positive	*Apteryx australis mantelli (North Island brown kiwi)*	scaffold1034	96	78	1.00E-06
3009	CCCGGGAGCTTTTCGGGCTGTGCATGTCGA AGCGGTGCCGGGGAGCAGCCGGCTGCATTT TGCGATAGGTGGAAATC	min	0.141	Negative	*Low/poor conservation*	[redacted]			

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
3153	CCCGGGCTCCTCACCCCCGCAGCCGGGCAACGCCTCTGCCGCACCCGCAGGGATGGGCCCTGGGCTGGGAAGGGGTC	min	0.168	Positive	*Low/poor conservation*	[redacted]			
3291	CCCGGGGCGGCTTTAGCACAGGGAGAGCTCGGCAGGGAGCGATAATAACATCTCCCCCGGCCTCTGCGCAGGAATAT	min	0.097	Positive	*Low/poor conservation*				
3436	CCCGGGCACGCAGCTCCTCCACCCTGGGGACACCCGCATGGCAGCCAGCCCGGCACGCTGCCCAGCGCCGGCGCGGC	min	0.075	Negative	*Low/poor conservation*				
3550	CCCGGGGGAGGGACGGGCACGTGCCCTCCGGGGAGCCGCACCGACTTCCCACGCTGCACACAAATGGACTCTACAAA	min	0.081	Positive	*Aptenodytes forsteri (Emperor penguin)*	CBFA2/RUNX1 translocation partner 2 (mRNA)	100	94	6.00E-23
3819	CCCGGGCTCTGGTTTTGTTCCTCAGCCTCTGCTTTTTTTCCTGGGGCTCTGGTTTTGTTCCTCGGGCTCTGGTTTTG	min	0.084	Positive	*Low/poor conservation*	[redacted]			
3839	CCCGGGCCGGACACAGCCGGTTCCAGCCGGCTCCAACGGCCCCAGCGCAGGGCACAGCTGAGCCCCTCAGCCGCGGG	min	0.067	Positive	*Calidris pugnax (Ruff)*	U1 small nuclear ribonucleoprotein C-like, transcript variant X1 (mRNA)	100	96	1.00E-25
3904	CCCGGGGGTGGGGTGGGACCCCTGGTTGTGGGGCCGCAGCTGCCCCCGGCAGCACCGAGGGCAGGGGTTGGGGACCCC	min	0.136	Positive	*Low/poor conservation*	[redacted]			
3953	CCCGGGGGCTGTGCTGGCAGCGTGCCGACGGCACAGTAGGATCCAAGCTGTGCTCCGGGACGAGCGGGGCTGGGGCC	min	0.082	Negative	*Low/poor conservation*				
4052	CCCGGGGAGCTGTCCCACCACCCCTCGCCATTGAGTCCATCAATGTCAAAAGTCTCCCGTGGCACGTGGGACTCCAG	min	0.049	Negative	*Seriola dumerili (Greater amberjack)*	SSC5D-like, transcript variant X1 (mRNA)	63	82	0.027

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
4295	CCCGGGCTGCCGGGCAGCACCTGCGGCTG CCGAAGGGAGACGGCCTTTGGTGGCGGGG AAGGGGCAGGAGGGCTGCA	min	0.178	Positive	*Low/poor conservation*				
4336	CCCGGGCTGCCGCAGAAGCACCTAGGGCA GACTGGTGATGAGCAGAATGGAGCAGAGGG GCGGGGAAGAGCGGGCTC	min	0.181	Negative	*Low/poor conservation*				
4422	CCCGGGACGTGCAGCGTTGGTTAACGCGCA CGGGTGAGTCAGGCCGGCGCAGACGAGAG GCACCGACCTGACCTGCC	min	0.059	Positive	*Low/poor conservation*				
4592	CCCGGGATGCTGGCCGGGGGACAGGGTGA CGAACCTCCCGCGACAGCGCGGCGCGAGG CTCGCCAGCCCCTGGCCTT	min	0.154	Negative	*Low/poor conservation*				
4608	CCCGGGGGCAGCTGGACCGTGGGTCTCCC CATGGGTGCTCCGCCGGGTCCCCACGGGTG AGGGGGCAGCCCTGGGGC	min	0.167	Negative	*Low/poor conservation*				
4733	CCCGGGGATGCCCTGCCCGTCCGGCCCCG TGTCCCCTCCTGCGGTGCCCGGATCTGGCC CCGTGCCGCAGGGAGGGG	min	0.004	Negative	*Columba livia (Rock pigeon)*	XKR7 (mRNA)	100	90	5.00E-18
5062	CCCGGGTCAGGAAGGAGATGCTGGAGGGC AGAAGCAACGTCCGTGCCAGGCCCCTGGCT CCGCGCCCCAGCCCAGCT	min	0.029	Negative	*Low/poor conservation*				
5424	CCCGGGGAACAAAACCAGAGGCCAAGGAAC AAAACCAGAGCCCTAGTAAGAAAACCAGAGC CCGAGTAACAAAACCA	min	0.049	Negative	*Pogona vitticeps (Central bearded dragon)*	FER1L5 (mRNA)	90	81	1.00E-05
8243	CCCGGGAGAGGCAGCCGCAGCCGCGCACA GCCCTCGCCCCAGCCCAAGCTATCCGCCAG CTGCTTTTGTAACATGGC	min	0.074	Positive	*Gavia stellata (Red-throated loon)*	SHC4 (mRNA)	94	89	6.00E-17
10174	CCCGGGGCTCTGGTTTTGTTATTCAGGCTCTG GTTTTGTTCCCGCGGGATCTGGTTTTCTCCC GCGGGCTCTGGTTTT	min	0.061	Positive	*Low/poor conservation*				
10344	CCCGGGGAATATAACCAGGGCCCGAGGAAC TAAACCAGAGCTCCAGGAACAAAACCAGAG CCCGAGGAACAAAACCA	min	0.015	Positive	*Low/poor conservation*				
12384	CCCGGGGAACAAAACCAGAGGCCGAGGAAC AAAGCCAGAGCCCGAGGAAGAAAACCAGAG	min	0.033	Negative	*Low/poor conservation*				

CpG	Sequence (Illumina NextSeq 77 bp)*	Lambda (λ) cut-off group	Unadjusted R^2	DNAm direction	BLAST species	BLAST region	Query cover (%)	Ident. (%)	E value
	CCCGCTGAACTCAAGAC								
12449	CCCGGGCACCCTGGCCCTGAAACAGCGCCTGCTGCTCTCCAAACACGTTCCCCTCCCTCCCACGTCCTGCCCAGGAG	min	0.063	Positive	*Apteryx australis mantelli (North Island brown kiwi)*	scaffold1015	75	88	7.00E-10
19826	CCCGGGGAACAAAACCAGAGGCTGAGGAACAAAGCCAGAGCCCTAGTAAGAAAACCAGAGCCCGAAGAACAAAACCA	min	0.053	Negative	*Diachasma alloeum (parasitic wasp)*	enolase-phosphatase E1-like (mRNA)	76	82	6.00E-04
25958	CCCGGGAAGAAAACGGTTGCCTGAGGAACAAACCAGAGCTCAAGGAACTCAACGCGAGCCCGAGGAAGAAAATCAG	min	0.174	Positive	*Low/poor conservation*				

Supplementary Table 4. DNA methylation levels of replicate samples.

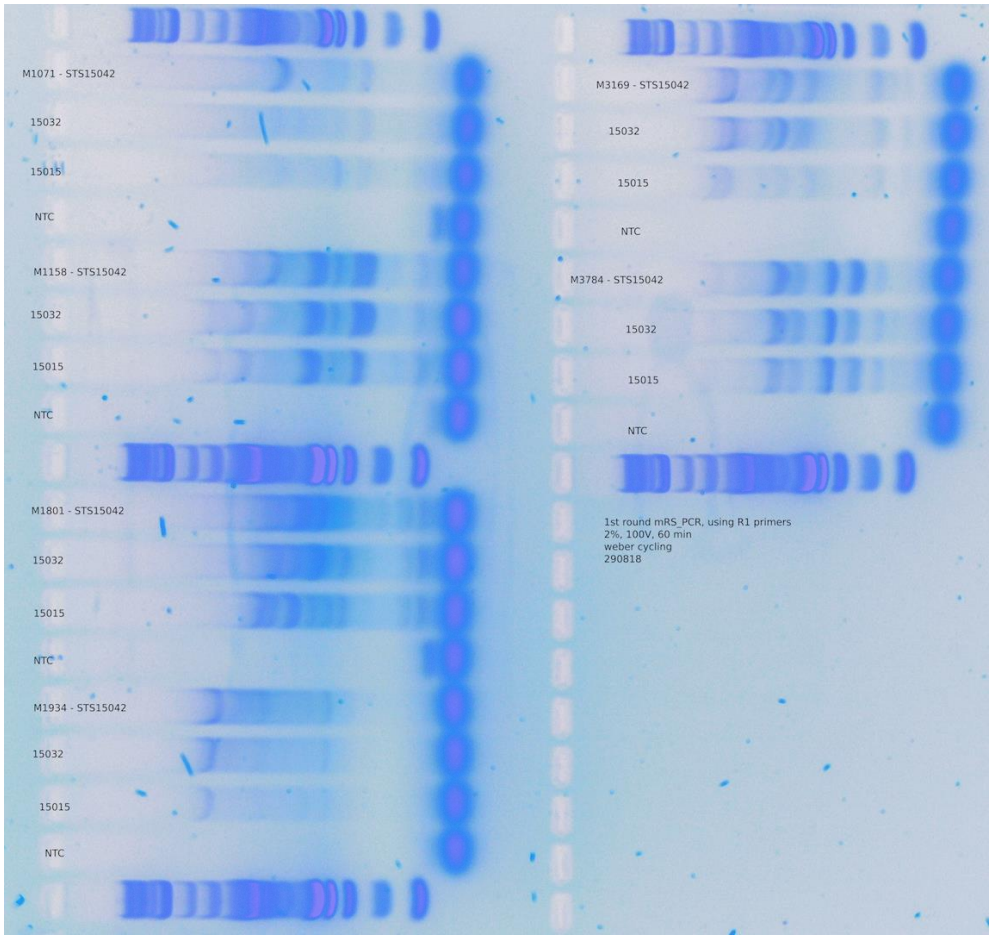
				CpG site							
UID	**Run**	**Type**	**Adjustment**	**M1071**	**M1158**	**M1801**	**M1934**	**M2083**	**M3169**	**M3784**	
15037	1	within_run	NA	0.505	0.153	0.773	0.208	0.667	0.246	0.824	
15037	1	within_run	NA	0.360	0.153	0.648	0.061	0.618	0.348	0.913	
16021	1	within_run	NA	0.350	0.153	0.604	0.083	0.636	0.217	1.000	
16021	1	within_run	NA	0.362	0.094	0.771	0.115	0.809	0.246	0.897	
15091	2	within_run	NA	0.279	0.233	0.632	0.061	0.684	0.156	0.700	
15091	2	within_run	NA	0.342	0.409	0.591	0.100	0.739	0.200	0.821	
16043	2	within_run	NA	0.481	0.069	0.634	0.085	0.489	0.130	0.872	
16043	2	within_run	NA	0.308	0.086	0.584	0.103	0.552	0.240	0.767	
17038	2	within_run	NA	0.360	0.037	0.612	0.061	0.622	0.227	0.807	
17038	2	within_run	NA	0.265	0.058	0.553	0.040	0.405	0.300	0.650	
15046	1	between_run	No	0.415	0.255	0.679	0.033	0.675	0.525	0.906	
15046	2	between_run	Pre	0.191	0.111	0.400	0.068	0.577	0.304	0.700	
15046	2	between_run	Post	0.342	0.174	0.549	0.083	0.514	0.470	0.840	
15057	1	between_run	No	0.434	0.258	0.750	0.081	0.556	0.143	0.821	
15057	2	between_run	Pre	0.292	0.250	0.640	0.080	0.784	0.050	0.730	
15057	2	between_run	Post	0.443	0.313	0.789	0.095	0.721	0.217	0.870	
15059	1	between_run	No	0.362	0.063	0.515	0.091	0.242	0.250	0.936	
15059	2	between_run	Pre	0.275	0.025	0.457	0.013	0.300	0.064	0.813	
15059	2	between_run	Post	0.426	0.089	0.606	0.028	0.237	0.231	0.953	
											Mean all CpGs (%)
Within run (mean absolute difference %)			-	11.91	7.13	7.90	5.18	11.13	7.58	11.34	8.65
Between run (mean absolute difference %)			Pre	15.12	6.31	14.91	3.79	16.32	15.72	14.90	11.29
Between run (mean absolute difference %)			Post	4.86	5.40	8.63	4.21	11.04	4.93	4.39	6.83

Note

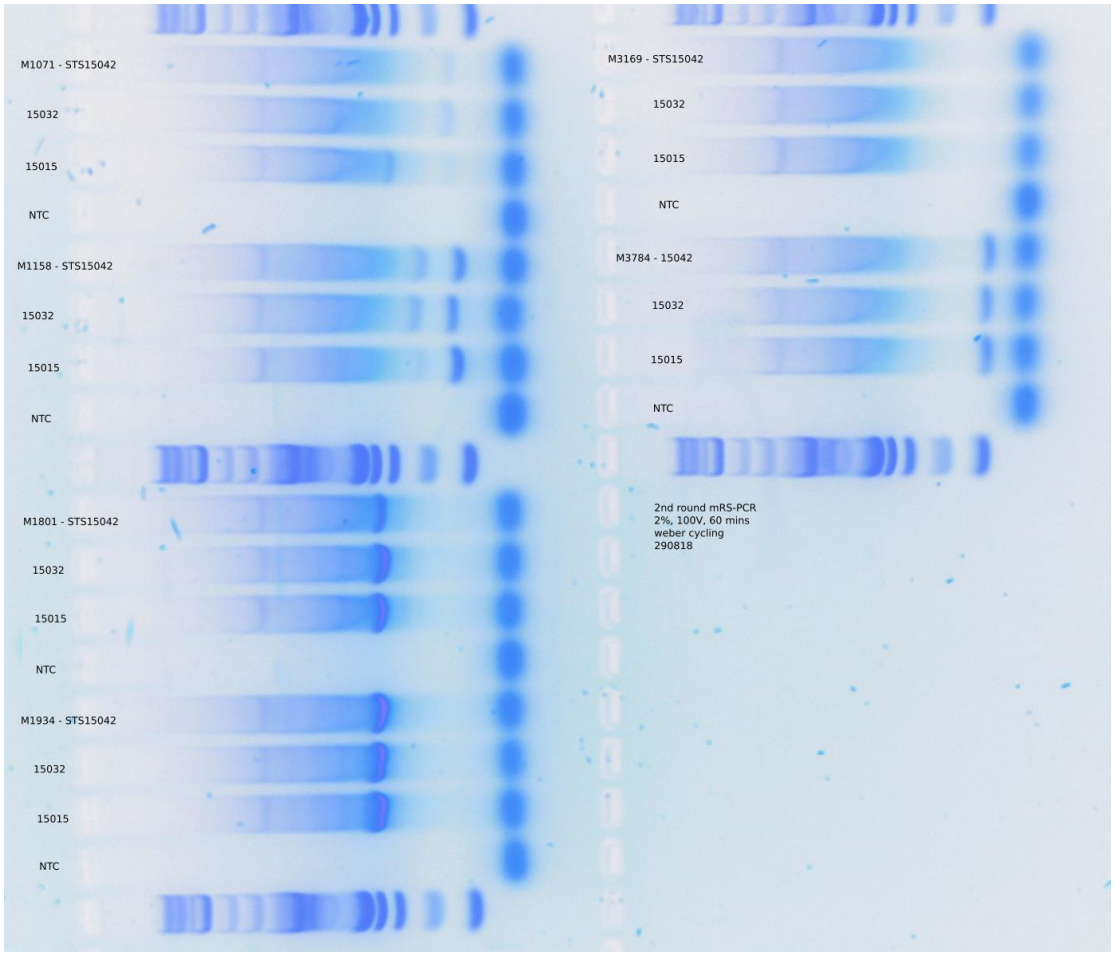
Yellow highlight indicates imputed values and are not included for calculations.

Supplementary Material

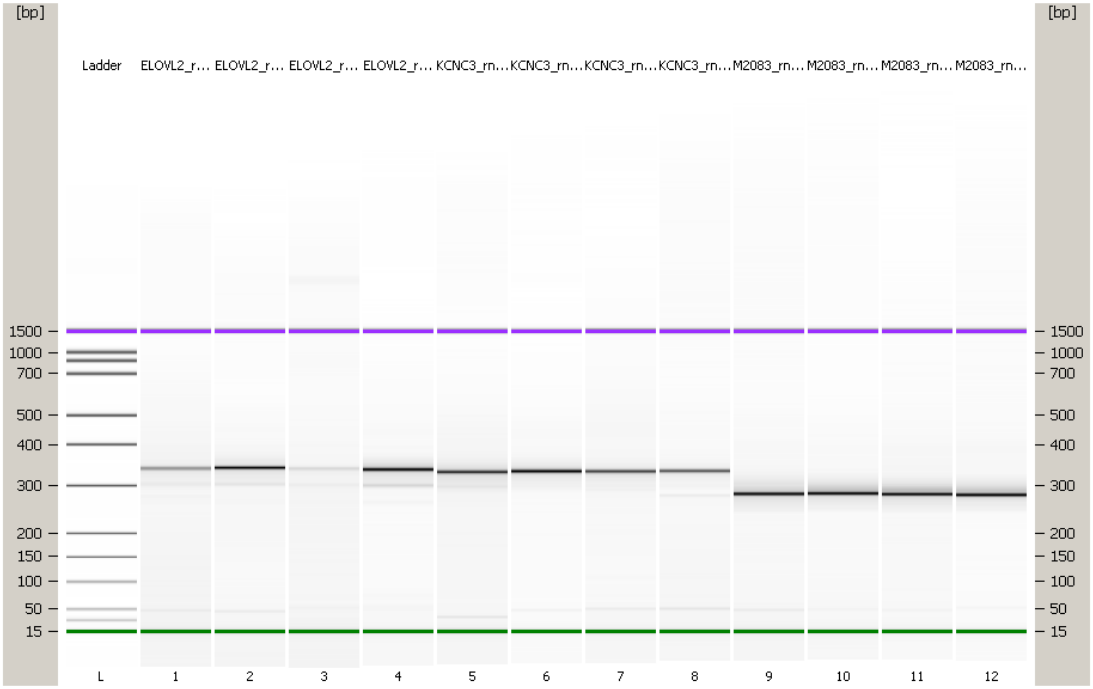
Chapter 5



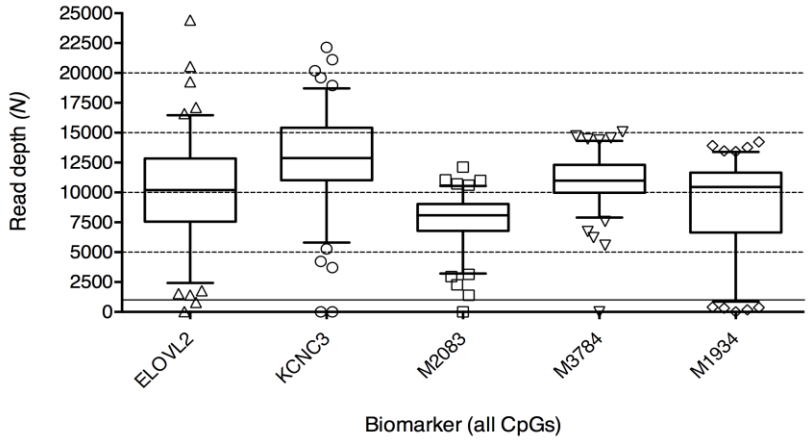
Supplementary Figure 1. Example gel electrophoresis of first round multiplexed restriction site PCR (mRS-PCR). This gel shown a test of six amplicons that encompass an age-related CpG site previously identified using the DREAM method (M1071, M1158, M1801, M1934, M3169, M3784). The first round mRS-PCR is run with four restriction site primers and the first specific primer. Each amplicon was run with three DNA samples and a no template control (NTC). The gel was run at 2%, 100V for 60 minutes. Laddering seen here is not uncommon for the first round.

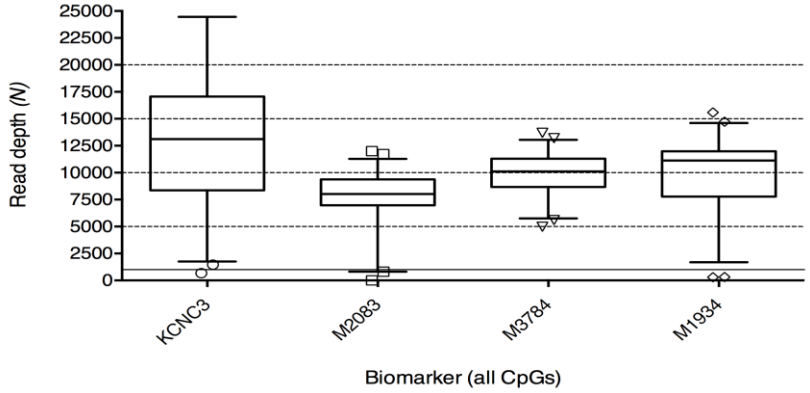


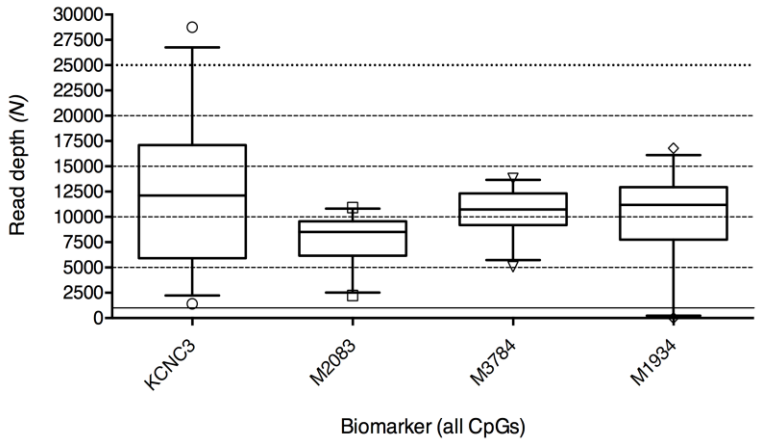
Supplementary Figure 2. Example gel electrophoresis of second round multiplexed restriction site PCR (mRS-PCR). This gel shown a test of six amplicons that encompass an age-related CpG site previously identified using the DREAM method (M1071, M1158, M1801, M1934, M3169, M3784). The second round mRS-PCR is again run with four restriction site primers and the second specific primer. Each amplicon was run with three DNA samples and a no template control (NTC). The gel was run at 2%, 100V for 60 minutes. In the second round, a single product is expected as seen for M1801, M1934, M1158 and M3784.



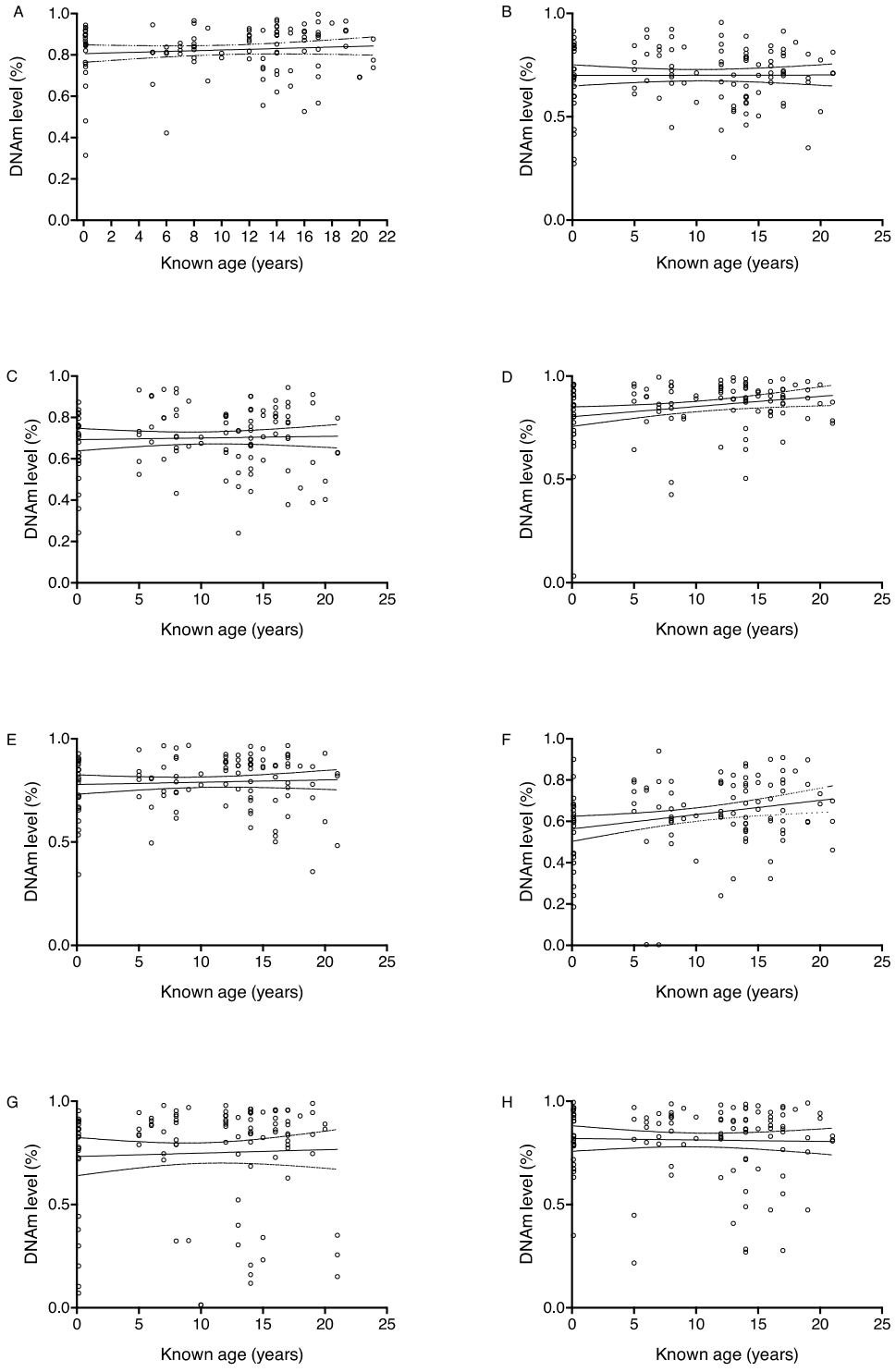
Supplementary Figure 3. Bioanalyzer quantification of amplicons following barcode attachment. An example of high quality barcode attachment is shown for amplicons *ELOVL2*, *KCNC3* and *M2083*. These samples can then be pooled without cross-contamination.



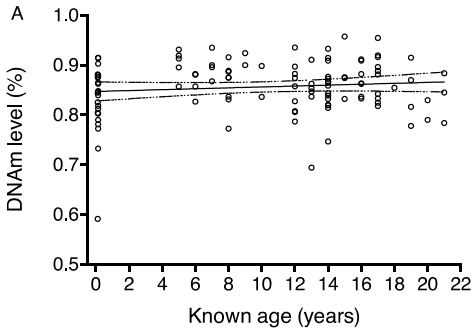


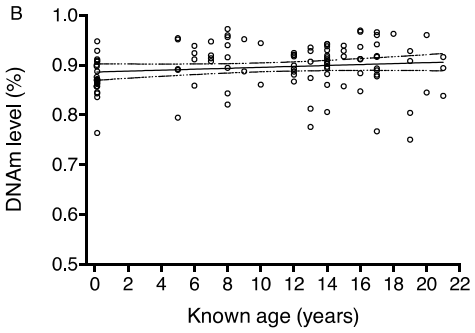


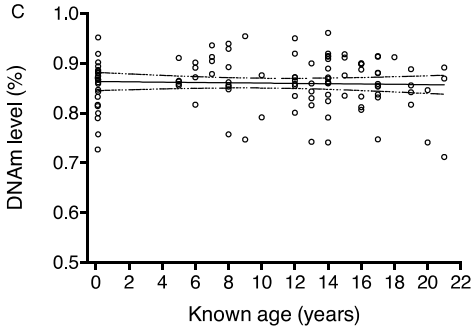
Supplementary Figure 4A – C. Box and whisker plots of sample read depths for MiSeq run. A. The read depths for known age individuals (*N* = 111, including chicks and technical replicates) included in the MiSeq run. **B.** The read depths for minimum age samples (*N* = 55). **C.** The read depths for unknown age samples (*N* = 24). The solid line on the Y-axis indicates a read depth of 1000x. DNAm scores calculated using less than 100x reads were not included for further analysis. All plots show 5 – 95 percentile.

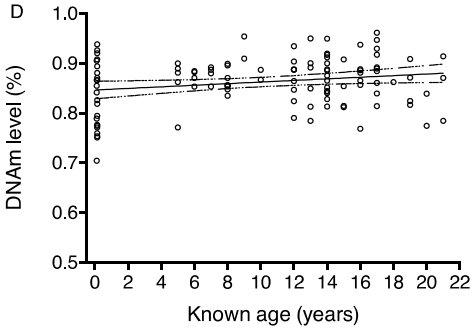


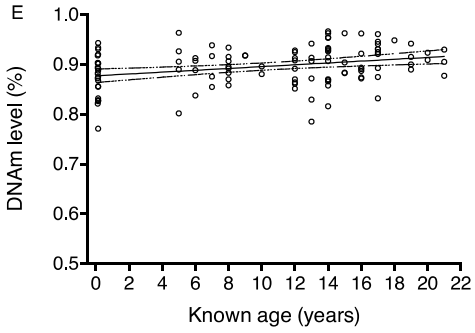
Supplementary Figure 5A – H. Linear regression for *ELOVL2* CpG sites. DNAm levels for eight CpG sites in *ELOVL2* were checked using simple linear regression against known-age (from leg bands) for the Short-tailed shearwater.

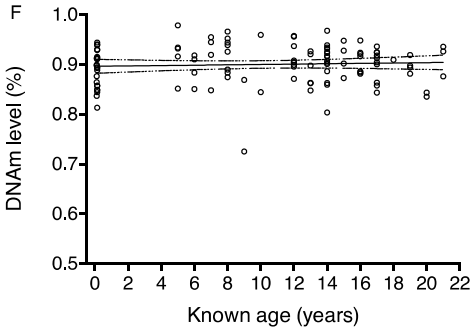

A
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

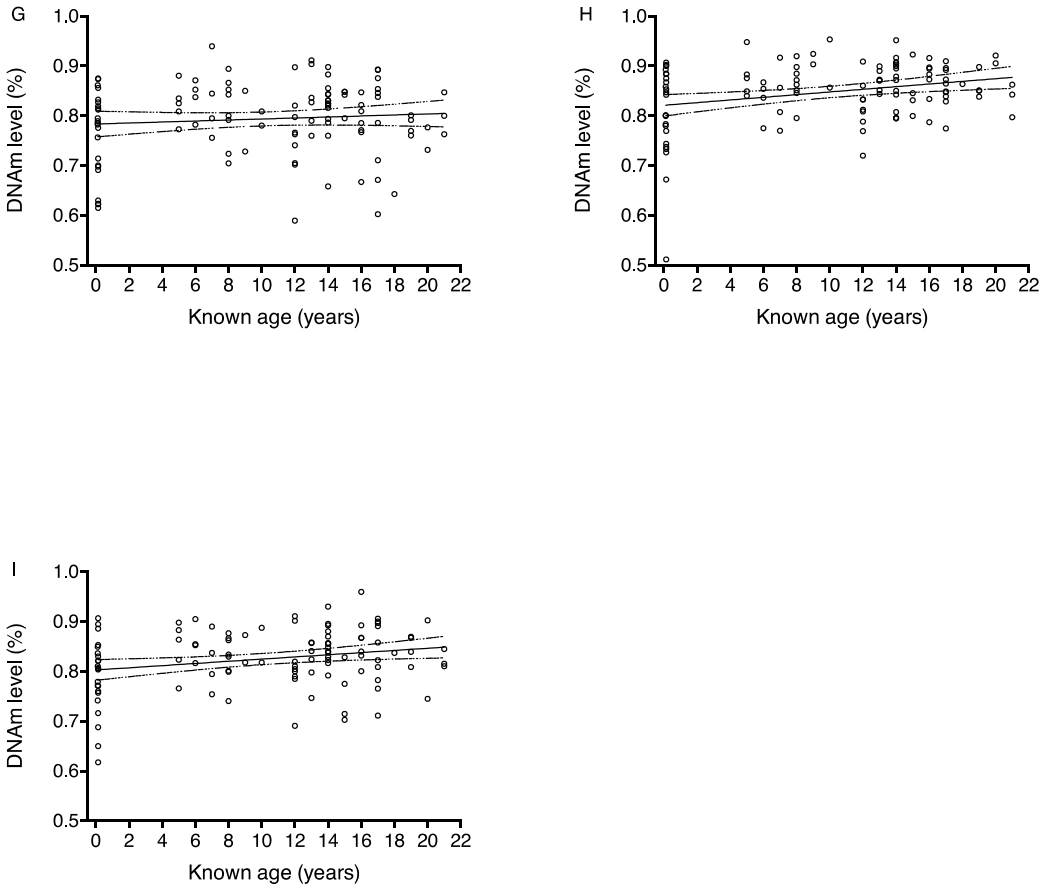

B
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)


C
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

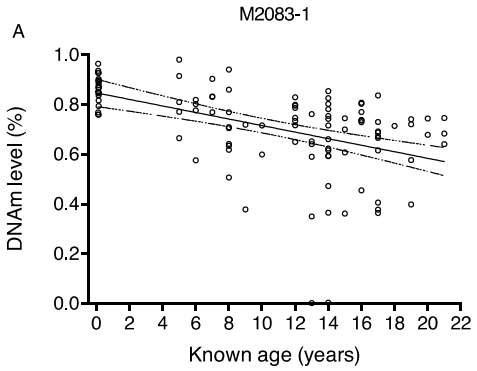

D
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

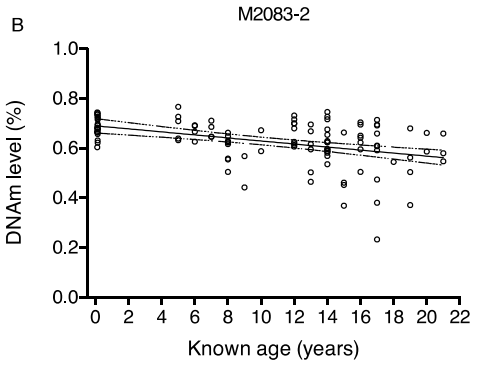

E
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

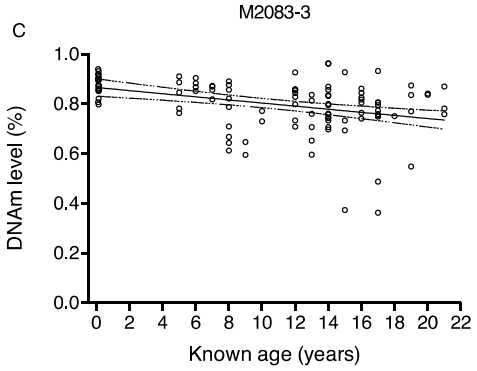

F
DNAm level (%)
1.0
0.9
0.8
0.7
0.6
0.5
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

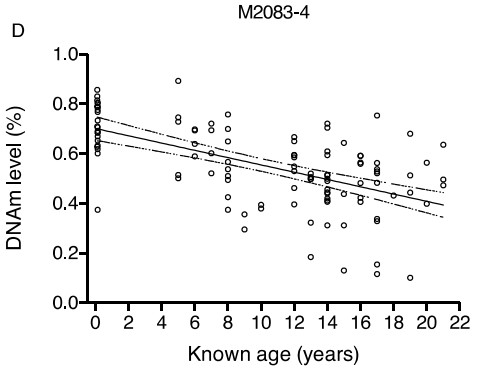


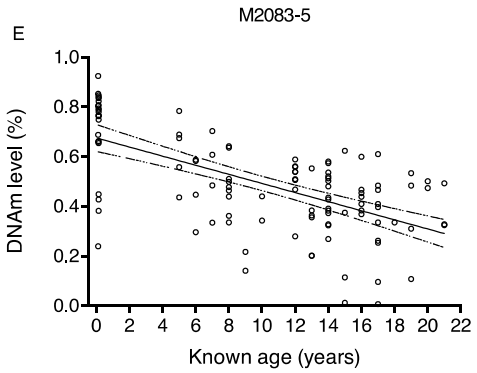
Supplementary Figure 6A – I. Linear regression for *KCNC3* CpG sites. DNAm levels for nine CpG sites in *KCNC3* were checked using simple linear regression against known-age (from leg bands) for the Short-tailed shearwater.

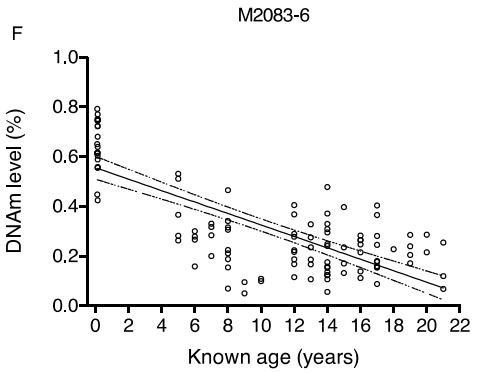

A
M2083-1
DNAm level (%)
Known age (years)

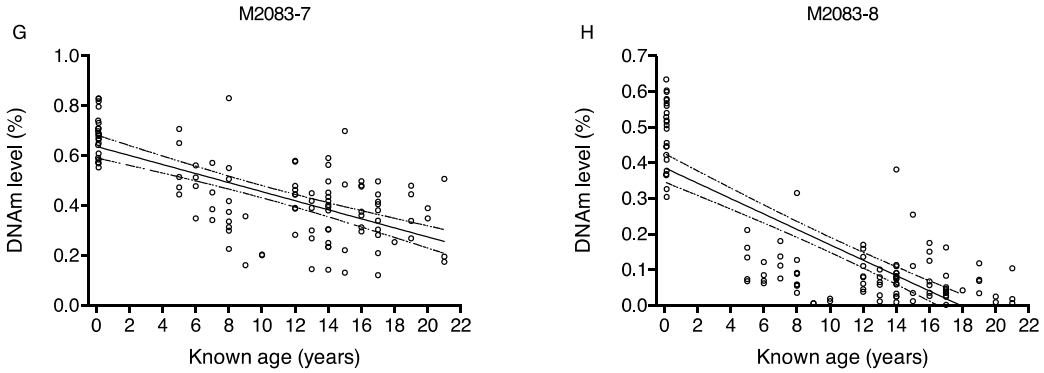

B
M2083-2
DNAm level (%)
Known age (years)


C
M2083-3
DNAm level (%)
Known age (years)

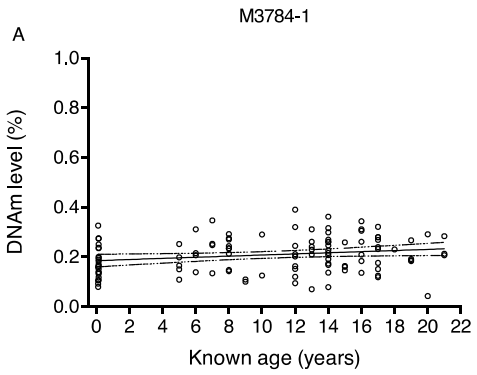

D
M2083-4
DNAm level (%)
Known age (years)

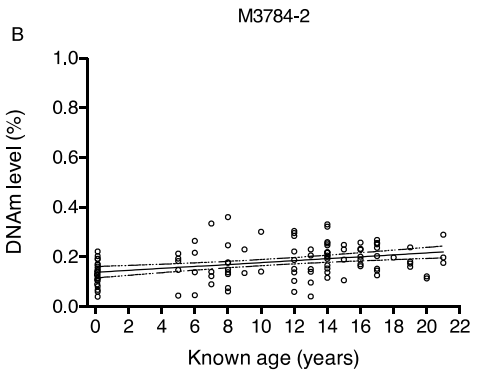

E
M2083-5
DNAm level (%)
Known age (years)

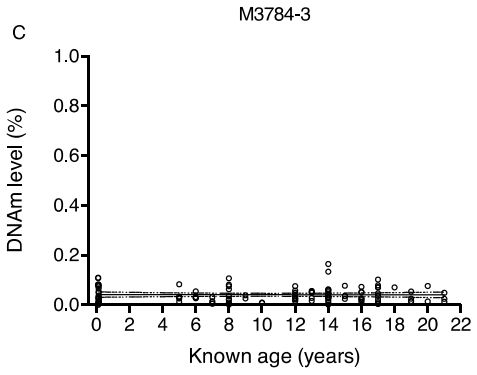

F
M2083-6
DNAm level (%)
Known age (years)

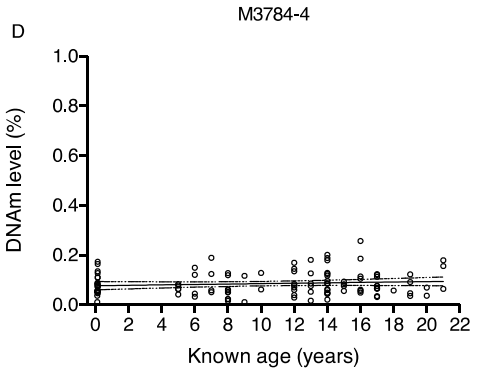


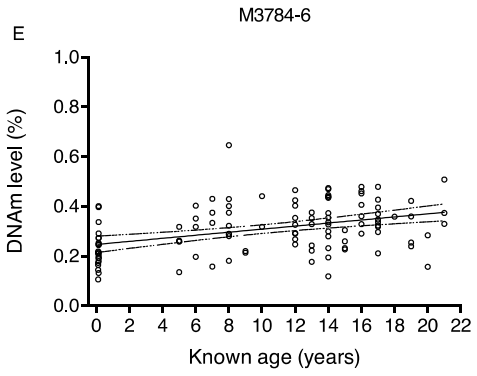
Supplementary Figure 7A – H. Linear regression for M2083 CpG sites. DNAm levels for eight CpG sites in the M2083 amplicon were checked using simple linear regression against known-age (from leg bands) for the Short-tailed shearwater. M2083-7 was the original aDMP identified using DREAM. CpGs are shown in 5' to 3' sequence.

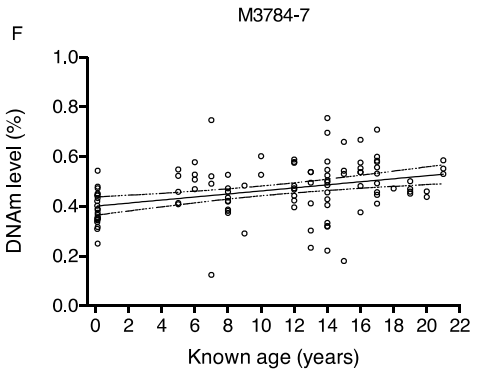

A
M3784-1
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

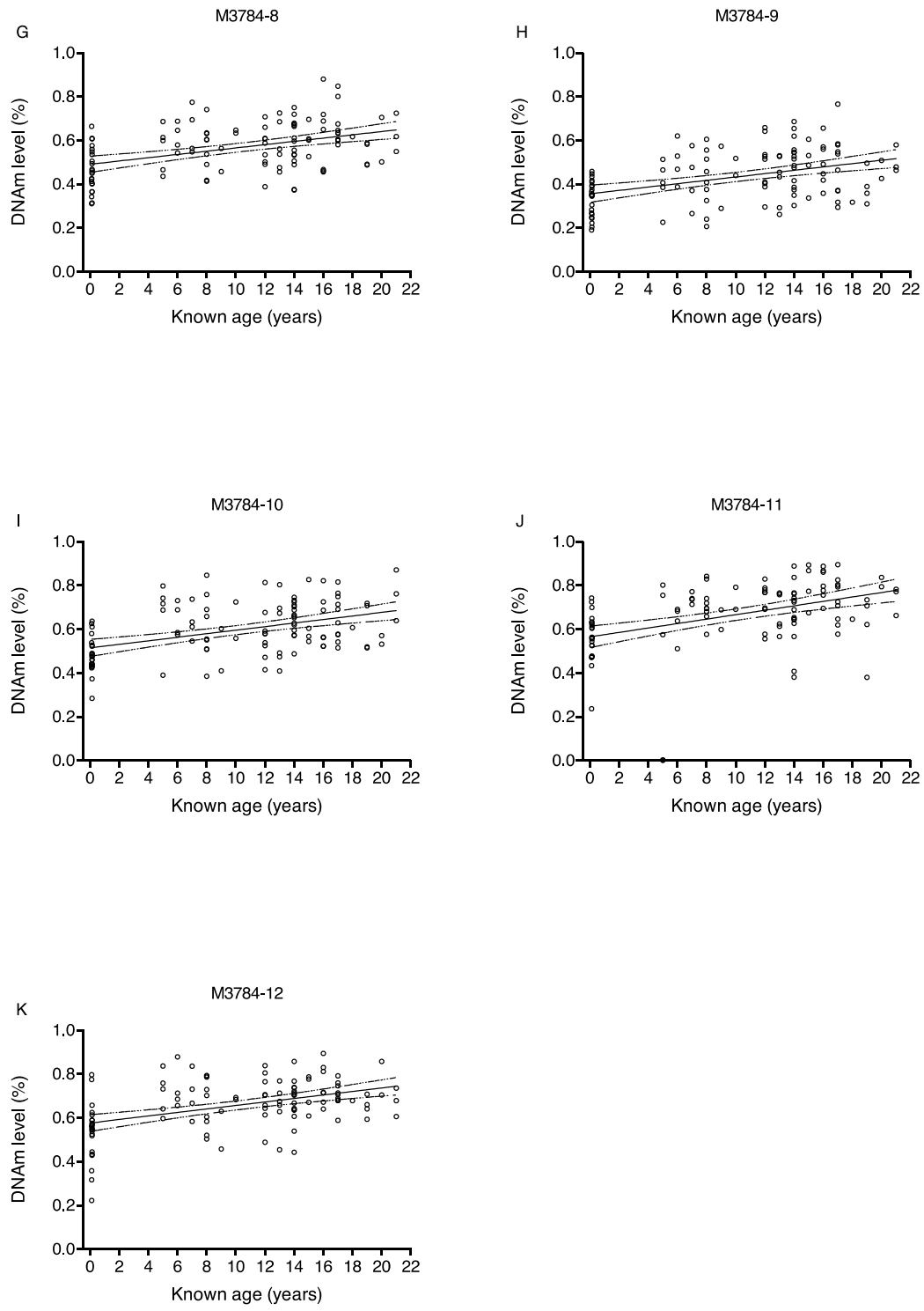

B
M3784-2
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)


C
M3784-3
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

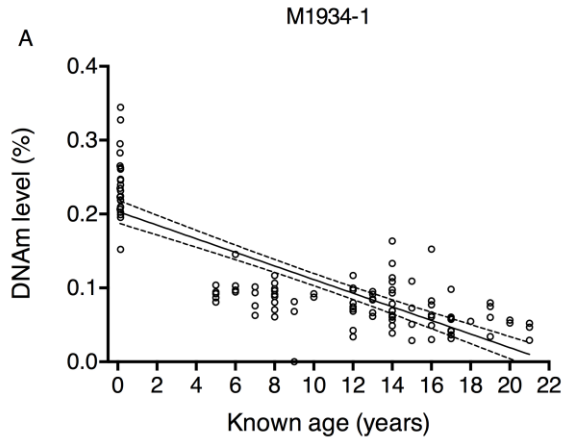

D
M3784-4
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)


E
M3784-6
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)

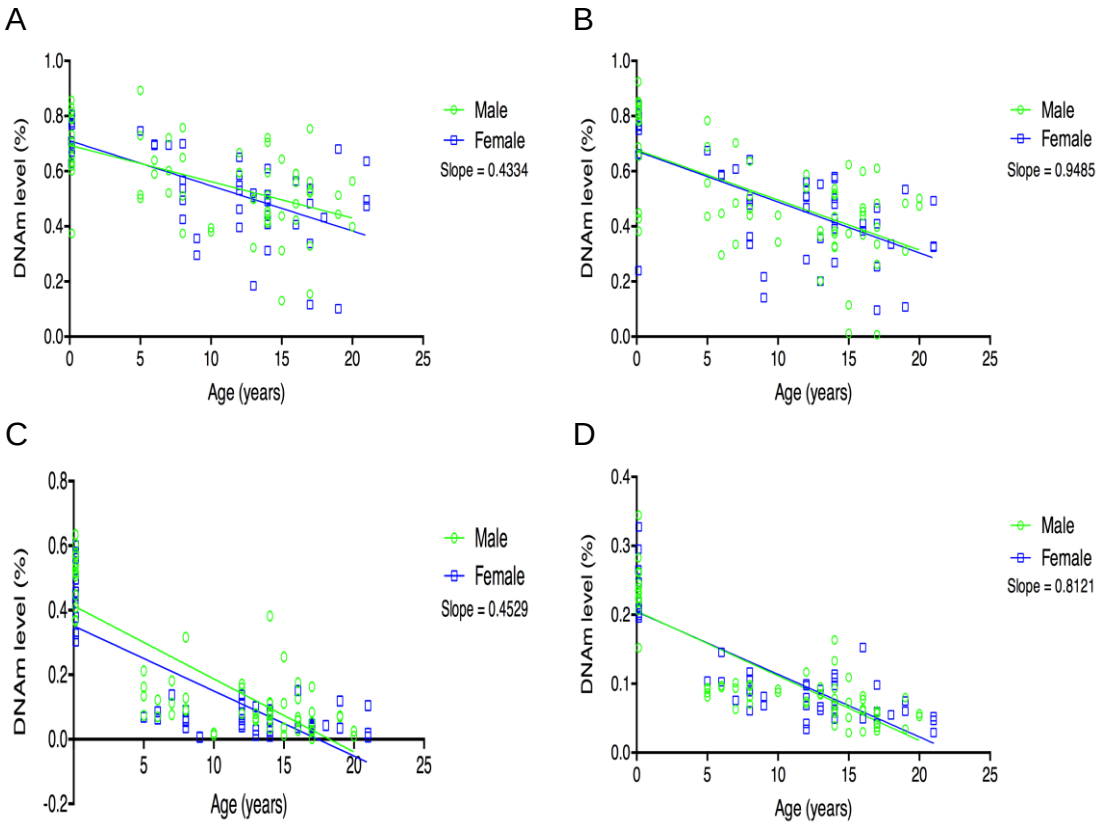

F
M3784-7
DNAm level (%)
1.0
0.8
0.6
0.4
0.2
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Known age (years)



Supplementary Figure 8A – K. Linear regression for M3784 CpG sites. DNAm levels for eleven CpG sites in the M3784 amplicon were checked using simple linear regression against known-age (from leg bands) for the Short-tailed shearwater. M3784-11 was the original aDMP identified using DREAM. CpGs are shown in 5' to 3' sequence.

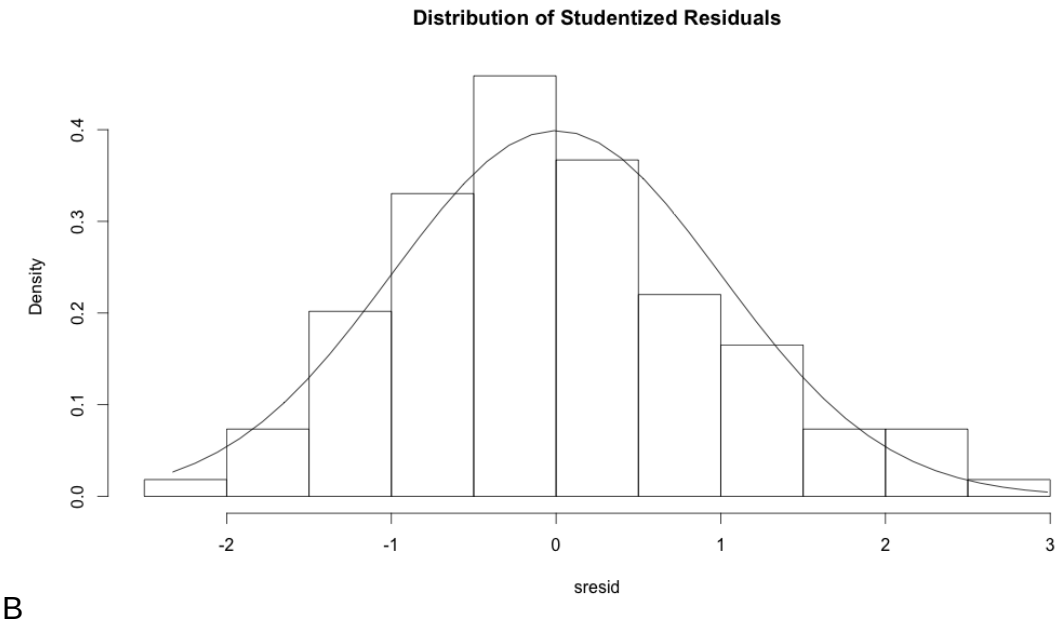


Supplementary Figure 9. Linear regression for M1934. DNAm levels for a single CpG site in the M1934 amplicon was checked using simple linear regression against known-age (from leg bands) for the Short-tailed shearwater. This was the original aDMP identified using DREAM.

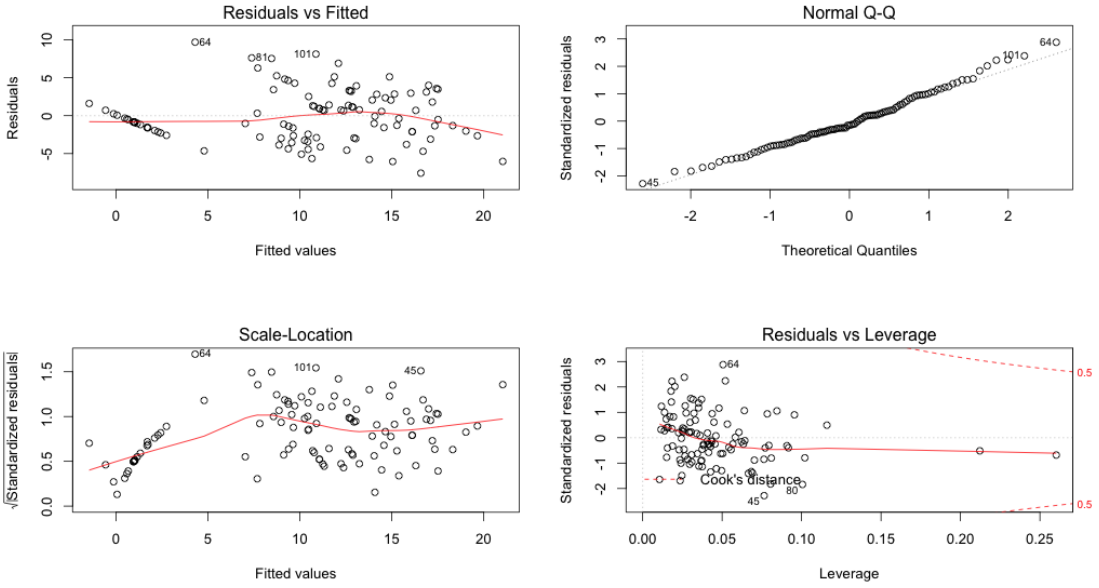


Supplementary Figure 10A – D. Linear regression for sites used in shearwater epigenetic age assay (SEAA). No significant sex differences were observed for the four age-related CpG sites used in the age assay (**A.** M2083-4, **B.** M2083-5, **C.** M2083-8 and **D.** M1934-1). Males and females are shown in green and blue respectively.

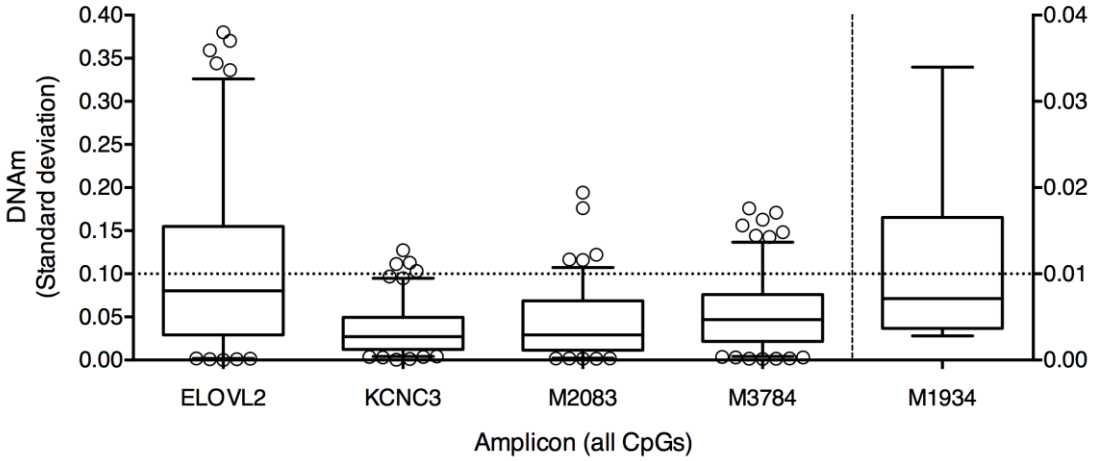
A



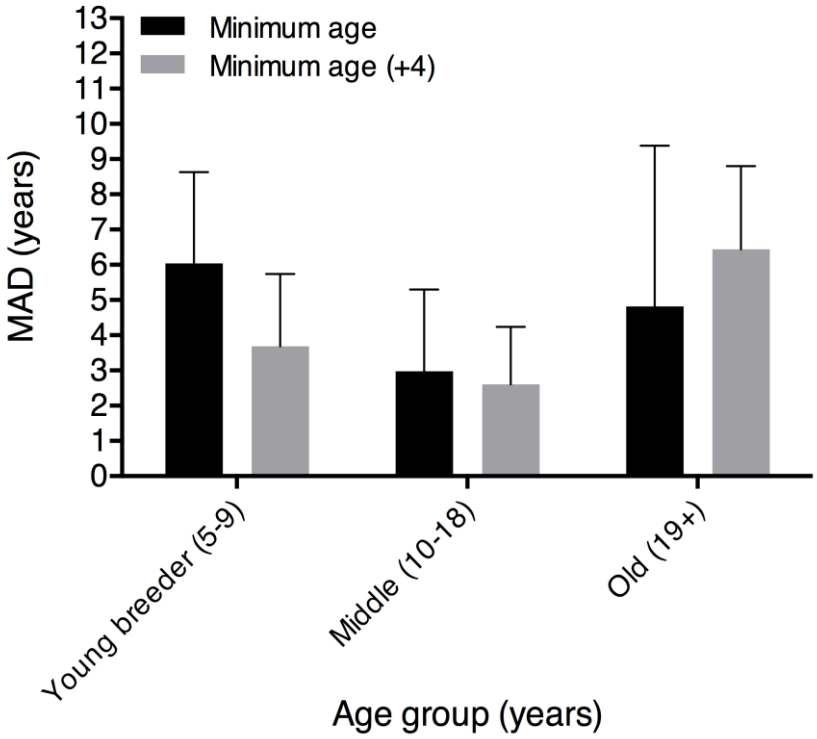
B



Supplementary Figure 11A – B. Diagnostics for Shearwater epigenetic age assay (SEAA) four age-related differentially methylated position (aDMP) model. A. The distribution of studentised residuals shows that the data is normally distributed in addition to standard multiple linear regression diagnostic plots (**B**).

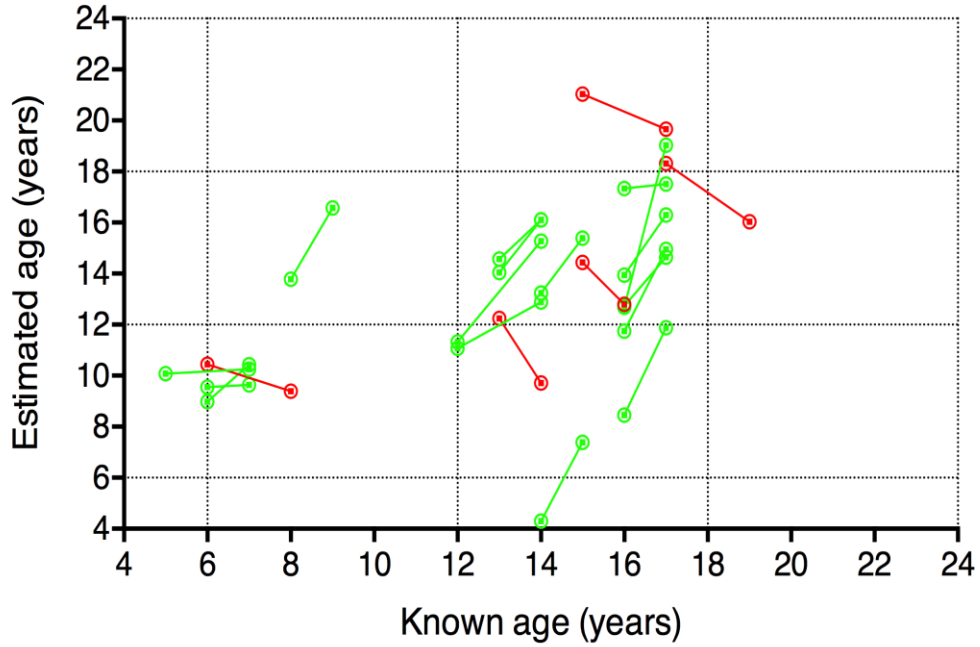


Supplementary Figure 12. Technical replicate DNAm standard deviation (SD). The DNAm level standard deviation between within-run technical replicates (*N* = 14) is shown for each of the assessed amplicons (5 – 95 percentile). ELOVL2 had the highest difference between identical samples (mean = 10.3%). The y-axis dashed line indicates a standard deviation in DNAm level of 0.10 (10%). M1934 is plotted on the right hand side y-axis. Overall mean SD was 4.9%. Data is from known-age samples.

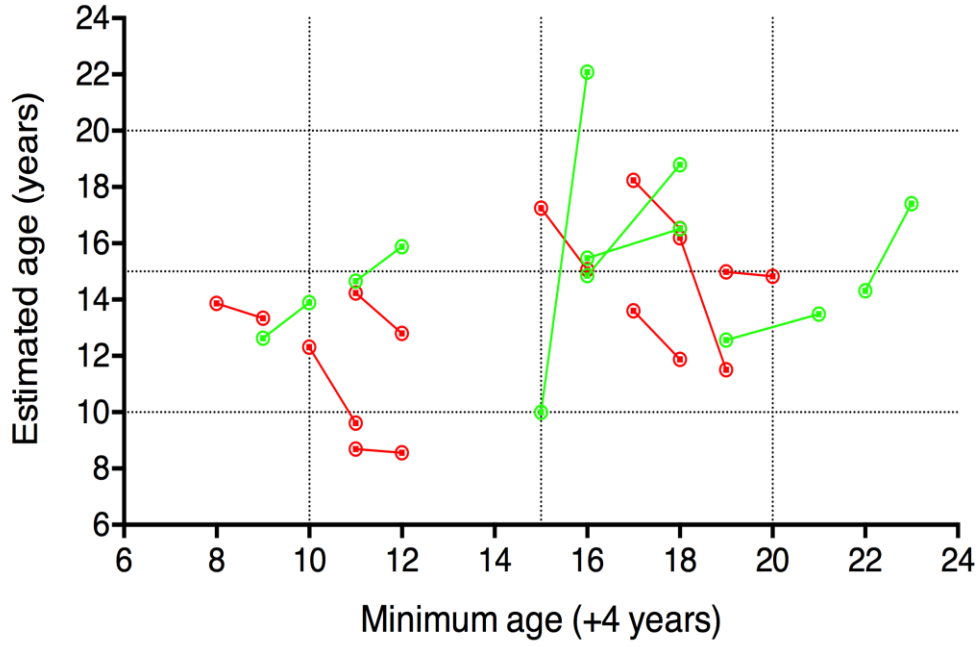


Supplementary Figure 13. Age-class grouped mean absolute deviation (MAD) for minimum-age samples. Determined by the absolute difference between the estimated and minimum age. Mean MADs are shown for each of the described age groups.

A



B



Supplementary Figure 14A – B. Longitudinal data for resighted individuals. Pairs of DNAm age estimates are shown for **A.** individuals in the known-age cohort (*N* = 21 pairs) and **B.** individuals in the minimum-age cohort (*N* = 16 pairs). Green points indicate the expected increase in estimated age given the time period between resights, either one or two years. Red points indicate a negative change between resights.

Supplementary Table 1 Sex information for samples included in bisulphite converted NGS validation

Group	Male	Female
Known age	41	35
Chicks	12	8
Minimum age	27	28
Unknown age	19	5

Supplementary Table 2 Primers for multiplexed restriction-site PCR (mRS-PCR).

Restriction site / biomarker	Direction	Sequence	Tm (°C)	bp
RSO-Bam	F	TAATACGACTCACTATAGGGAGANNNNNNNNNNGGATCC	61.3	39
RSO-Eco	F	TAATACGACTCACTATAGGGAGANNNNNNNNNGAATTC	59.2	39
RSO-Sau	F	TAATACGACTCACTATAGGGAGANNNNNNNNNGATC	58.9	37
RSO-Taq	F	TAATACGACTCACTATAGGGAGANNNNNNNNTCGA	58.9	37
*M1071_R1	R	TGTTTAGTTCCTCTGGCTCTGG	60.8	22
*M1071_R2	R	TCCTGGAGCTCTGGTTCAGT	60	20
M1158_R1	R	ACCGGCACGTCCCTCTC	62.3	17
M1158_R2	R	GGCTTCGCGCTTTTGG	61.6	16
M1801_R1	R	AGACGCCGGCAGAGC	59.7	15
M1801_R2	R	GCTACATTTCCCGGCATCT	60.1	19
M1934_R1	R	TTTGTGGTTTTTGGAAGTACTGA	58.7	23
M1934_R2	R	GAGATTTCATTGCTGCTTATTCAA	59.8	24
M2083_R1	R	CCTTGGAGATCCTGCCTTC	59.7	19
M2083_R2	R	TCCTCCTTCTCCTCCTCCTC	59.9	20
*M3169_R1	R	GCACGAACTGCTTGTAGAGC	58.8	20
*M3169_R2	R	AGCGGCGAGAGCTGC	60.1	15
M3784_F1	F	CTAAGCTGCCAGGAAGGAAA	59.6	20
M3784_R1	R	GCTCCGGCTGCTCTTTG	60.8	17
M3784_R2	R	GCATCTCTCAGCCCAGATGT	60.4	20

* Note: These primers did not generate a single band appropriate for sequencing.

Supplementary Table 3 Multiplexed restriction site PCR (mRS-PCR) cycling

conditions

Cycles (x)	Temperature (°C)	Time
1	95	10 min
40*	95	1 min
	55	2 min
	72	3 min
1	72	10 min
∞	4	∞

*Cycles were reduced to 20x for agarose band stab isolated product amplification.

Supplementary Table 4 BrightDye® Terminator sequencing cycling conditions

Cycles (x)	Temperature (°C)	Time
1	96	3 mins
30	96	10 sec
	50	5 sec
	60	2.30 mins
∞	4	∞

Supplementary Table 5 ZymoTaq bisulphite converted PCR cycling conditions

Cycles (x)	Temperature (°C)	Time
1	95	10 mins
45	95	30 secs
	60*	30 secs
	72	30 secs
1	72	7 mins
∞	4	∞

* Temperature is optimised for each amplicon as appropriate.

Supplementary Table 6 mRS-PCR generated genomic sequence BLASTn results (all organisms)

Restriction site / biomarker	Gene	Region	Organism	Total score	Query cover (%)	E value	Ident (%)	CpGs (N)	Sequence length (bp)	GC content (%)
M1158	None	-	-	-	-	-	-	30	245	74.3
M1801	LLGL1	mRNA	*Corvus brachyrhynchos (American crow)*	223	46	4.00E-56	92	28	381	69.3
M1934	Uncharacterised	ncRNA	*Anas platyrynchos (mallard)*	260	54	2.00E-65	87	2	357	40.9
M1934*	RUNX1	RefSeqGene	*Homo sapiens (human)*	134	34	2.00E-27	82			
M2083	TMBIM1	mRNA	*Serinus canaria (Atlantic canary)*	100	59	3.00E-17	71	13	339	69.9
M3784	Uncharacterised	Scaffold	*Apteryx australis mantelli (North Island Brown Kiwi)*	314	82	1.00E-57	73	43	595	68.2

*Duplicates are included (where possible) when the top match was uncharacterised in bird.

Supplementary Table 7 All CpG sites screened for age-related DNAm in *Ardenna tenuirostris*

Biomarker	CpG	R^2 (all samples)	Correlation direction	*P**	Reference
ELOVL2	1	0.0091	NA	*P* > 0.05	Garagnani et al. (2012), Polanowski et al. (2014), Slieker et al. (2018).
	2	0.0000	NA	*P* > 0.05	
	3	0.0011	NA	*P* > 0.05	
	4	0.0548	Positive	*P* = 0.014	
	5	0.0030	NA	*P* > 0.05	
	6	0.0644	Positive	*P* = 0.008	
	7	0.0017	NA	*P* > 0.05	
	8	0.0007	NA	*P* > 0.05	
KCNC3	1	0.0120	NA	*P* > 0.05	Rakyan et al . (2010), Koch et al. (2011), De Paoli-Iseppi et al. (2017).
	2	0.0174	NA	*P* > 0.05	
	3	0.0016	NA	*P* > 0.05	
	4	0.0427	Positive	*P* = 0.031	
	5	0.0904	Positive	*P* = 0.002	
	6	0.0037	NA	*P* > 0.05	
	10	0.0080	NA	*P* > 0.05	
	11	0.0771	Positive	*P* = 0.004	
	12	0.0561	Positive	*P* = 0.013	
M2083	1	0.2396	Negative	*P* < 0.0001	De Paoli-Iseppi et al. (2018).
	2	0.1908	Negative	*P* < 0.0001	
	3	0.1444	Negative	*P* < 0.0001	
	4	0.3332	Negative	*P* < 0.0001	
	5	0.3785	Negative	*P* < 0.0001	
	6	0.5705	Negative	*P* < 0.0001	
	7	0.4541	Negative	*P* < 0.0001	
	8	0.6110	Negative	*P* < 0.0001	
M3784	1	0.0404	Positive	*P* = 0.036	De Paoli-Iseppi et al. (2018).
	2	0.1325	Positive	*P* < 0.0001	
	3	0.0001	NA	*P* > 0.05	
	4	0.0125	NA	*P* > 0.05	
	6	0.1587	Positive	*P* < 0.0001	
	7	0.1293	Positive	*P* = 0.0001	
	8	0.1783	Positive	*P* < 0.0001	
	9	0.1681	Positive	*P* < 0.0001	
	10	0.1874	Positive	*P* < 0.0001	
	11	0.1848	Positive	*P* < 0.0001	
	12	0.1890	Positive	*P* < 0.0001	
M1934	1	0.6509	Negative	*P* < 0.0001	De Paoli-Iseppi et al. (2018).

*Unadjusted P values shown. Bonferroni-Holm corrected alpha value is used for significance (α = .0014).

Supplementary Table 8 DNAm level standard deviation for technical replicates in each MiSeq amplicon

Amplicon	N*	Minimum	Maximum	Mean
ELOVL2	112	0.0001	0.3800	0.1035
KCNC3	126	0.0004	0.1273	0.0343
M2083	112	0.0015	0.1943	0.0412
M3789	154	0.0012	0.1760	0.0545
M1934	14	0.0028	0.0340	0.0116
				0.0490

* N indicates the number of DNAm values recorded for all CpGs in an amplicon

References

Acker P, Robert A, Bourget R, Colas B (2014) Heterogeneity of reproductive age increases the viability of semelparous populations. *Functional Ecology* **28**, 458-468.

Aguilar JRd, Westerdahl H, Puente JMdl*, et al.* (2016) MHC-I provides both quantitative resistance and susceptibility to blood parasites in blue tits in the wild. *Journal of Avian Biology* **47**, 669-677.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Amendt J, Richards C, Campobasso C, Zehner R, Hall M (2011) Forensic entomology: applications and limitations. *Forensic science, medicine, and pathology* **7**, 379-392.

Andraszek K, Gryzińska M, Wójcik E, Knaga S, Smalec E (2014) Age-dependent change in the morphology of nucleoli and methylation of genes of the nucleolar organizer region in Japanese quail (*Coturnix japonica*) model (Temminck and Schlegel, 1849)(Galliformes: Aves). *Folia Biologica* **62**, 293-300.

Arima H, Ohnishi N (2006) Usefulness of avian buccal cells for molecular sexing. *Ornithological Science* **5**, 139-143.

Baccarelli A, Wright RO, Bollati V*, et al.* (2009) Rapid DNA methylation changes after exposure to traffic particles. *American journal of respiratory and critical care medicine* **179**, 572-578.

Barrès R, Osler ME, Yan J*, et al.* (2009) Non-CpG methylation of the PGC-1α promoter through DNMT3B controls mitochondrial density. *Cell metabolism* **10**, 189-198.

Barres R, Yan J, Egan B*, et al.* (2012) Acute exercise remodels promoter methylation in human skeletal muscle. *Cell metabolism* **15**, 405-411.

Bauer CM, Graham JL, Abolins-Abols M*, et al.* (2018) Chronological and Biological Age Predict Seasonal Reproductive Timing: An Investigation of Clutch Initiation and Telomeres in Birds of Known Age. *The American Naturalist* **191**, 777-782.

Baylis SM, Sunnucks P, Clarke R (2018) A model for first-estimates of species-specific, age-specific mortality from centralized band-recovery databases. *Ecosphere* **9**, e02136.

Beeton NJ, McMahon CR, Williamson GJ*, et al.* (2015) Using the spatial population abundance dynamics engine for conservation management. *Methods in Ecology and Evolution* **6**, 1407-1416.

Beissinger SR, Peery MZ (2007) Reconstructing the historic demography of an endangered seabird. *Ecology* **88**, 296-305.

Beissinger SR, Westphal MI (1998) On the use of demographic models of population viability in endangered species management. *The Journal of wildlife management*, 821-841.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. *Molecular Ecology Resources* **9**, 1279-1301.

Bekaert B, Kamalandua A, Zapico SC, Van de Voorde W, Decorte R (2015) Improved age determination of blood and teeth samples using a selected set of DNA methylation markers. *Epigenetics* **10**, 922-930.

Belinsky SA, Palmisano WA, Gilliland FD*, et al.* (2002) Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Research* **62**, 2370-2377.

Bell JT, Tsai P-C, Yang T-P*, et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet* **8**, e1002629.

Bello N, Francino O, Sánchez A (2001) Isolation of genomic DNA from feathers. *Journal of Veterinary diagnostic investigation* **13**, 162-164.

Benhamed M, Herbig U, Ye T, Dejean A, Bischof O (2012) Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nature cell biology* **14**, 266-275.

Bentz AB, Sirman AE, Wada H, Navara KJ, Hood WR (2016) Relationship between maternal environment and DNA methylation patterns of estrogen receptor alpha in wild Eastern Bluebird (Sialia sialis) nestlings: a pilot study. *Ecology and Evolution* **6**, 4741-4752.
Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A (2009) An operational definition of epigenetics. *Genes & development* **23**, 781-783.
Bernstein BE, Meissner A, Lander ES (2007) The mammalian epigenome. *Cell* **128**, 669-681.
Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes & development* **16**, 6-21.
Bird A (2007) Perceptions of epigenetics. *Nature* **447**, 396-398.
Bize P, Criscuolo F, Metcalfe NB, Nasir L, Monaghan P (2009) Telomere dynamics rather than age predict life expectancy in the wild. *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 1679-1683.
Bjorksten J (1968) The crosslinkage theory of aging. *Journal of the American Geriatrics Society* **16**, 408-427.
Bocklandt S, Lin W, Sehl ME*, et al.* (2011) Epigenetic predictor of age.
Boks MP, Derks EM, Weisenberger DJ*, et al.* (2009) The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One* **4**, e6767.
Bollati V, Baccarelli A (2010) Environmental epigenetics. *Heredity* **105**, 105.
Bosnjak J, Stevanov-Pavlovic M, Vucicevic M*, et al.* (2013) Feasibility of non-invasive molecular method for sexing of parrots. *Pakistan Journal of Zoology* **45**.
Bouwmeester MC, Ruiter S, Lommelaars T*, et al.* (2016) Zebrafish embryos as a screen for DNA methylation modifications after compound exposure. *Toxicology and Applied Pharmacology* **291**, 84-96.
Bradley J, Cox J, Nicholson L*, et al.* (2000a) Parental influence upon the provisioning schedules of nestling Short-tailed Shearwaters Puffinus tenuirostris. *Journal of Avian Biology* **31**, 522-526.
Bradley J, Skira I, Wooller R (1991) A long-term study of Short-tailed Shearwaters (*Puffinus tenuirostris*) on Fisher Island, Australia. *Ibis* **133**, 55-61.
Bradley J, Wooller R, Skira I (2000b) Intermittent breeding in the short-tailed shearwater Puffinus tenuirostris. *Journal of Animal Ecology* **69**, 639-650.
Bradley J, Wooller R, Skira I, Serventy D (1989) Age-dependent survival of breeding Short-tailed Shearwaters (*Puffinus tenuirostris*). *The Journal of Animal Ecology*, 175-188.
Bradley RJ, Safran RJ (2014) Conceptual Revision and Synthesis of Proximate Factors Associated with Age-Related Improvement in Reproduction. *Ethology* **120**, 411-426.
Branco MR, Ficz G, Reik W (2012) Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Reviews Genetics* **13**, 7-13.
Bravington MV, Grewe PM, Davies CR (2016) Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications* **7**, 13162.
Bravington MV, Jarman SN, Skaug HJ (2014) Antarctic Blue Whale surveys: augmenting via genetics for close-kin and ordinal age.
Brooke M (2013) *The manx shearwater* A&C Black.
Brooke MdL, Bonnaud E, Dilley B*, et al.* (2018) Seabird population changes following mammal eradications on islands. *Animal Conservation* **21**, 3-12.
Brown MB, Brown CR (2009) Blood sampling reduces annual survival in cliff swallows (Petrochelidon pyrrhonota). *The Auk* **126**, 853-861.
Buckmeier DL, Irwin ER, Betsill RK, Prentice JA (2002) Validity of otoliths and pectoral spines for estimating ages of channel catfish. *North American Journal of Fisheries Management* **22**, 934-942.
Burger J, Schreiber E, Hamer KC (2001) Breeding biology, life histories, and life history–environment interactions in seabirds. In: *Biology of marine birds*, pp. 230-275. CRC Press.
Campana S (2001) Accuracy, precision and quality control in age determination, including a review of the use and abuse of age validation methods. *Journal of Fish Biology* **59**, 197-242.

Caracappa S, Pisciotta A, Persichetti M*, et al.* (2016) Nonmodal scutes patterns in the Loggerhead Sea Turtle (Caretta caretta): a possible epigenetic effect? *Canadian Journal of Zoology* **94**, 379-383.

Cavalcante RG, Patil S, Park Y, Rozek LS, Sartor MA (2017) Integrating DNA Methylation and Hydroxymethylation Data with the Mint Pipeline. *Cancer Research* **77**, e27-e30.

Cerchiara JA, Risques RA, Prunkard D*, et al.* (2017) Telomeres shorten and then lengthen before fledging in Magellanic penguins (*Spheniscus magellanicus*). *Aging (Albany NY)* **9**, 487.

Chambers LE, Patterson T, Hobday AJ*, et al.* (2015) Determining trends and environmental drivers from long-term marine mammal and seabird data: examples from Southern Australia. *Regional environmental change* **15**, 197-209.

Chaney Jr RC, Blemings KP, Bonner J, Klandorf H (2003) Pentosidine as a measure of chronological age in wild birds. *The Auk* **120**, 394-399.

Charpentier M, Tung J, Altmann J, Alberts S (2008) Age at maturity in wild baboons: genetic, environmental and demographic influences. *Molecular Ecology* **17**, 2026-2040.

Christensen BC, Houseman EA, Marsit CJ*, et al.* (2009) Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* **5**, e1000602.

Christiansen L, Lenart A, Tan Q*, et al.* (2016) DNA methylation age is associated with mortality in a longitudinal Danish twin study. *Aging Cell* **15**, 149-154.

Christidis L, Boles W (2008) *Systematics and taxonomy of Australian birds* CSIRO Publishing.

Christman JK (2002) 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-5495.

Clark SJ, Lee HJ, Smallwood SA, Kelsey G, Reik W (2016) Single-cell epigenomics: powerful new methods for understanding gene regulation and cell identity. *Genome Biology* **17**, 72.

Clay TA, Pearmain EJ, McGill RA, Manica A, Phillips RA (2018) Age-related variation in non-breeding foraging behaviour and carry-over effects on fitness in an extremely long-lived bird. *Functional Ecology*.

Clutton-Brock T, Sheldon BC (2010) Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution* **25**, 562-573.

Cobb JS, Wahle RA (1994) Early life history and recruitment processes of clawed lobsters. *Crustaceana* **67**, 1-25.

Cooey CK, Fallon JA, Avery ML*, et al.* (2010) Refinement of biomarker pentosidine methodology for use on aging birds. *Human-Wildlife Interactions* **4**, 304-314.

Cook PE, Hugo LE, Iturbe-Ormaetxe I*, et al.* (2006) The use of transcriptional profiles to predict adult mosquito age under field conditions. *Proceedings of the National Academy of Sciences* **103**, 18060-18065.

Cortés V, García-Barcelona S, González-Solís J (2018) Sex-and age-biased mortality of three shearwater species in longline fisheries of the Mediterranean. *Marine Ecology Progress Series* **588**, 229-241.

Dale VH, Beyeler SC (2001) Challenges in the development and use of ecological indicators. *Ecological indicators* **1**, 3-10.

De Magalhaes J, Costa J (2009) A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology* **22**, 1770-1774.

De Paoli-Iseppi R, Deagle BE, McMahon CR*, et al.* (2017a) Measuring animal age with DNA methylation: From humans to wild animals. *Frontiers in Genetics* **8**, 106.

De Paoli-Iseppi R, Polanowski AM, McMahon C*, et al.* (2017b) DNA methylation levels in candidate genes associated with chronological age in mammals are not conserved in a long-lived seabird. *PLoS One* **12**, e0189181.

De Paoli-Iseppi R, Deagle B, Polanowski A*, et al.* (2018) Age estimation in a long-lived seabird (Ardenna tenuirostris) using DNA methylation-based biomarkers. *Molecular Ecology Resources*.

Dolinoy DC, Huang D, Jirtle RL (2007) Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. *Proceedings of the National Academy of Sciences* **104**, 13056-13061.

Dorr BS, Stahl RS, Hanson-Dorr KC, Furcolow CA (2017) Using pentosidine and hydroxyproline to predict age and sex in an avian species. *Ecology and Evolution* **7**, 8999-9005.

Dunshea G, Duffield D, Gales N*, et al.* (2011) Telomeres as age markers in vertebrate molecular ecology. *Molecular Ecology Resources* **11**, 225-235.

Eaton MJ, Link WA (2011) Estimating age from recapture data: integrating incremental growth measures with ancillary data to infer age-at-length. *Ecological Applications* **21**, 2487-2497.

Eckalbar WL, Schlebusch SA, Mason MK*, et al.* (2016) Transcriptomic and epigenomic characterization of the developing bat wing. *Nature genetics*.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461.

Edgar RC, Flyvbjerg H (2015) Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**, 3476-3482.

Egan B, Carson BP, Garcia-Roves PM*, et al.* (2010) Exercise intensity-dependent regulation of peroxisome proliferator-activated receptor γ coactivator-1α mRNA abundance is associated with differential activation of upstream signalling kinases in human skeletal muscle. *The Journal of physiology* **588**, 1779-1790.

Egan B, Zierath JR (2013) Exercise metabolism and the molecular regulation of skeletal muscle adaptation. *Cell metabolism* **17**, 162-184.

Ehrich M, Nelson MR, Stanssens P*, et al.* (2005) Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proceedings of the National Academy of Sciences* **102**, 15785-15790.

Ekroos J, Öst M, Karell P, Jaatinen K, Kilpi M (2012) Philopatric predisposition to predation-induced ecological traps: habitat-dependent mortality of breeding eiders. *Oecologia* **170**, 979-986.

Ellegren H (1996) First gene on the avian W chromosome (CHD) provides a tag for universal sexing of non-ratite birds. *Proceedings of the Royal Society of London B: Biological Sciences* **263**, 1635-1641.

Elliott KH, Hare JF, Le Vaillant M*, et al.* (2015) Ageing gracefully: physiology but not behaviour declines with age in a diving seabird. *Functional Ecology* **29**, 219-228.

Ernande B, Clobert J, McCombie H, Boudry P (2003) Genetic polymorphism and trade-offs in the early life-history strategy of the Pacific oyster, *Crassostrea gigas* (Thunberg, 1795): a quantitative genetic study. *Journal of Evolutionary Biology* **16**, 399-414.

Escobar-Zepeda A, Vera-Ponce de León A, Sanchez-Flores A (2015) The road to metagenomics: from microbiology to DNA sequencing technologies and bioinformatics. *Frontiers in Genetics* **6**, 348.

Essington TE, Kitchell JF, Walters CJ (2001) The von Bertalanffy growth function, bioenergetics, and the consumption rates of fish. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 2129-2138.

Esteller M, Herman JG (2002) Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *The Journal of pathology* **196**, 1-7.

Fallon JA, Cochrane RL, Dorr B, Klandorf H (2006) Interspecies comparison of pentosidine accumulation and its correlation with age in birds. *The Auk* **123**, 870-876.

Farrar WL, Ruscetti FW, Young HA (1985) 5-Azacytidine treatment of a murine cytotoxic T cell line alters interferon-gamma gene induction by interleukin 2. *The Journal of Immunology* **135**, 1551-1554.

Faux CE, McInnes JC, Jarman SN (2014) High-throughput real-time PCR and melt curve analysis for sexing Southern Ocean seabirds using fecal samples. *Theriogenology* **81**, 870-874.
Fay R, Barbraud C, Delord K, Weimerskirch H (2016) Paternal but not maternal age influences early-life performance of offspring in a long-lived seabird. *Proc. R. Soc. B* **283**, 20152318.
Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics* **13**, 97-109.
Feng S, Cokus SJ, Zhang X*, et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences* **107**, 8689-8694.
Feng Y, Zhao L-Z, Hong L*, et al.* (2013) Alteration in methylation pattern of GATA-4 promoter region in vitamin A-deficient offspring's heart. *The Journal of nutritional biochemistry* **24**, 1373-1380.
Festa-Bianchet M, Blanchard P, Gaillard J-M, Hewison AM (2002) Tooth extraction is not an acceptable technique to age live ungulates. *Wildlife Society Bulletin* **30**, 282-283.
Fontana L, Partridge L, Longo VD (2010) Extending healthy life span—from yeast to humans. *Science* **328**, 321-326.
Fraga MF, Ballestar E, Paz MF*, et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10604-10609.
Frederiksen M, Moe B, Daunt F*, et al.* (2012) Multicolony tracking reveals the winter distribution of a pelagic seabird on an ocean basin scale. *Diversity and Distributions* **18**, 530-542.
Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **33**, 1.
Froy H, Phillips RA, Wood AG, Nussey DH, Lewis S (2013) Age-related variation in reproductive traits in the wandering albatross: evidence for terminal improvement following senescence. *Ecology Letters* **16**, 642-649.
Furness RW, Wade HM, Robbins AM, Masden EA (2012) Assessing the sensitivity of seabird populations to adverse effects from tidal stream turbines and wave energy devices. *ICES Journal of Marine Science* **69**, 1466-1479.
Gamelon M, Vriend SJ, Engen S*, et al.* (2019) Accounting for interspecific competition and age structure in demographic analyses of density dependence improves predictions of fluctuations in population size. *Ecology Letters*.
Gao X, Zhang Y, Breitling LP, Brenner H (2016) Relationship of tobacco smoking and smoking-related DNA methylation with epigenetic age acceleration. *Oncotarget* **7**, 46878.
Garagnani P, Bacalini MG, Pirazzini C*, et al.* (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* **11**, 1132-1134.
Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *Journal of Molecular Biology* **196**, 261-282.
Gaudet F, Hodgson JG, Eden A*, et al.* (2003) Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489-492.
Gayral P, Weinert L, Chiari Y*, et al.* (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Molecular Ecology Resources* **11**, 650-661.
Gianuca D, Phillips RA, Townley S, Votier SC (2017) Global patterns of sex-and age-specific variation in seabird bycatch. *Biological Conservation* **205**, 60-76.
Gill F, Donsker D (2013) IOC world bird list (v 3.5). Accessed online at http//www. worldbirdnames. org.
Glaze CM, Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. *The Journal of neuroscience* **26**, 991-1005.
Goyns MH (2002) Genes, telomeres and mammalian ageing. *Mechanisms of Ageing and Development* **123**, 791-799.
Green RE, Langston RH, McCluskie A, Sutherland R, Wilson JD (2016) Lack of sound science in assessing wind farm impacts on seabirds. *Journal of Applied Ecology* **53**, 1635-1641.

Gregory MK, Geier MS, Gibson RA, James MJ (2013) Functional characterization of the chicken fatty acid elongases. *The Journal of nutrition* **143**, 12-16.
Gregory TR (2001) Animal genome size database. TR Gregory.
Grönniger E, Weber B, Heil O*, et al.* (2010) Aging and chronic sun exposure cause distinct epigenetic changes in human skin. *PLoS Genet* **6**, e1000971.
Gruber BL, Sorbi D, French DL*, et al.* (1996) Markedly elevated serum MMP-9 (gelatinase B) levels in rheumatoid arthritis: a potentially useful laboratory marker. *Clinical immunology and immunopathology* **78**, 161-171.
Gryzińska M, Andraszek K, Jocek G (2013) DNA methylation analysis of the gene CDKN2B in Gallus gallus (Chicken). *Folia Biologica* **61**, 165-171.
Gryzinska M, Blaszczak E, Strachecka A, Jezewska-Witkowska G (2013) Analysis of age-related global DNA methylation in chicken. *Biochemical Genetics* **51**, 554-563.
Gryzinska M, Jakubczak A, Listos P*, et al.* (2016) Association between body weight and age of dogs and global DNA methylation. *Medycyna Weterynaryjna* **72**, 64-67.
Gu H, Smith ZD, Bock C*, et al.* (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**, 468.
Gunn JS, Clear NP, Carter TI*, et al.* (2008) Age and growth in southern bluefin tuna, *Thunnus maccoyii* (Castelnau): direct estimation from otoliths, scales and vertebrae. *Fisheries Research* **92**, 207-220.
Hall ME, Nasir L, Daunt F*, et al.* (2004) Telomere loss in relation to age and early environment in long-lived birds. *Proceedings of the Royal Society of London B: Biological Sciences* **271**, 1571-1576.
Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT **41**, 95-98.
Hamano Y, Manabe S, Morimoto C, Fujimoto S, Tamaki K (2017) Forensic age prediction for saliva samples using methylation-sensitive high resolution melting: exploratory application for cigarette butts. *Scientific Reports* **7**, 10444.
Handel CM, Pajot LM, Talbot SL, Sage GK (2006) Use of buccal swabs for sampling DNA from nestling and adult birds. *Wildlife Society Bulletin* **34**, 1094-1100.
Hannum G, Guinney J, Zhao L*, et al.* (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular Cell* **49**, 359-367.
Harris M (1966) Age of return to the colony, age of breeding and adult survival of Manx Shearwaters. *Bird study* **13**, 84-95.
Harris MP (2014) Aging Atlantic Puffins Fratercula arctica in summer and winter. *Seabird* **27**, 22-40.
Harvey MG, Bonter DN, Stenzler LM, Lovette IJ (2006) A comparison of plucked feathers versus blood samples as DNA sources for molecular sexing. *Journal of Field Ornithology* **77**, 136-140.
Haussmann MF, Mauck RA (2008) Telomeres and longevity: testing an evolutionary hypothesis. *Molecular Biology and Evolution* **25**, 220-228.
Haussmann MF, Winkler DW, O'Reilly KM*, et al.* (2003) Telomeres shorten more slowly in long-lived birds and mammals than in short–lived ones. *Proceedings of the Royal Society of London B: Biological Sciences* **270**, 1387-1392.
Hayflick L (1985) Theories of biological aging. *Experimental Gerontology* **20**, 145-159.
Hayflick L (2007) Biological aging is no longer an unsolved problem. *Annals of the New York Academy of Sciences* **1100**, 1-13.
Head JA (2014) Patterns of DNA methylation in animals: an ecotoxicological perspective. *Integrative and comparative biology*, icu025.
Hernando-Herraez I, Prado-Martinez J, Garg P*, et al.* (2013) Dynamics of DNA methylation in recent human and great ape evolution. *PLoS Genet* **9**, e1003763.
Hilborn R, Walters CJ (1992) Quantitative fisheries stock assessment: choice, dynamics and uncertainty. *Reviews in Fish Biology and Fisheries* **2**, 177-178.

Hollander MC, Blumenthal GM, Dennis PA (2011) PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nature Reviews Cancer* **11**, 289-301.
Holliday R, Pugh JE (1996) DNA modification mechanisms and gene activity during development. *Cold Spring Harbor Monograph Archive* **32**, 639-645.
Horn T, Gemmell N, Robertson B, Bridges C (2008) Telomere length change in European sea bass (Dicentrarchus labrax). *Australian journal of zoology* **56**, 207-210.
Horvath S (2013) DNA methylation age of human tissues and cell types. *Genome Biology* **14**, R115.
Horvath S, Erhart W, Brosch M*, et al.* (2014) Obesity accelerates epigenetic aging of human liver. *Proceedings of the National Academy of Sciences* **111**, 15538-15543.
Horvath S, Raj K (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics*, 1.
Hu X, Sui X, Li L*, et al.* (2013a) Protocadherin 17 acts as a tumour suppressor inducing tumour cell apoptosis and autophagy, and is frequently methylated in gastric and colorectal cancers. *The Journal of pathology* **229**, 62-73.
Hu Y, Xu H, Li Z*, et al.* (2013b) Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broilers. *PLoS One* **8**, e56411.
Hunter CM, Caswell H (2005) Selective harvest of sooty shearwater chicks: effects on population dynamics and sustainability. *Journal of Animal Ecology* **74**, 589-600.
Iqbal M, Probert LL, Alhumadi NH, Klandorf H (1999) Protein glycosylation and advanced glycosylated endproducts (AGEs) accumulation: an avian solution? *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* **54**, B171-B176.
Issa JP (2000) CpG-island methylation in aging and cancer. *Current topics in microbiology and immunology* **249**, 101.
Ito G, Yoshimura K, Momoi Y (2017) Analysis of DNA methylation of potential age-related methylation sites in canine peripheral blood leukocytes. *Journal of Veterinary Medical Science* **79**, 745-750.
Ito H, Udono T, Hirata S, Inoue-Murayama M (2018) Estimation of chimpanzee age based on DNA methylation. *Scientific Reports* **8**, 9998.
IUCN (2012) *BirdLife International. 2012. Ardenna tenuirostris. The IUCN Red List of Threatened Species 2012: e.T22698216A40211770.* http://dx.doi.org/10.2305/IUCN.UK.2012-1.RLTS.T22698216A40211770.en
Iverson JB, Stahl RS, Furcolow C, Kraus F (2017) An evaluation of the use of pentosidine as a biomarker for ageing turtles. *Conservation physiology* **5**, cow076.
Jakubczak A, Listos P, Dudko P, Abramowicz K, Jeżewska-Witkowska G (2016) Association between body weight and age of dogs and global DNA methylation. *Medycyna Weterynaryjna* **72**, 64-67.
Jarman SN, Polanowski AM, Faux CE*, et al.* (2015) Molecular biomarkers for chronological age in animal ecology. *Molecular Ecology* **24**, 4826-4847.
Jazwinski SM, Kim S (2017) Metabolic and Genetic Markers of Biological Age. *Frontiers in Genetics* **8**.
Jelinek J, Liang S, Lu Y*, et al.* (2012) Conserved DNA methylation patterns in healthy blood cells and extensive changes in leukemia measured by a new quantitative technique. *Epigenetics* **7**, 1368-1378.
Jelinek J, Madzo J (2016) DREAM: A Simple Method for DNA Methylation Profiling by High-throughput Sequencing. *Chronic Myeloid Leukemia: Methods and Protocols*, 111-127.
Jennings S, Reynolds JD, Mills SC (1998) Life history correlates of responses to fisheries exploitation. *Proceedings of the Royal Society of London B: Biological Sciences* **265**, 333-339.
Jensen T, Pernasetti FM, Durrant B (2003) Conditions for rapid sex determination in 47 avian species by PCR of genomic DNA from blood, shell-membrane blood vessels, and feathers. *Zoo Biology* **22**, 561-571.
Jetz W, Thomas G, Joy J, Hartmann K, Mooers A (2012) The global diversity of birds in space and time. *Nature* **491**, 444-448.

Jones MJ, Goodman SJ, Kobor MS (2015) DNA methylation and healthy human aging. *Aging Cell*.
Jones OR, Scheuerlein A, Salguero-Gómez R*, et al.* (2014) Diversity of ageing across the tree of life. *Nature* **505**, 169-173.
Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**, 484-492.
Jones PA, Liang G (2009) Rethinking how DNA methylation patterns are maintained. *Nature Reviews Genetics* **10**, 805-811.
Jouventin P, Lequette B, Dobson FS (1999) Age-related mate choice in the wandering albatross. *Animal behaviour* **57**, 1099-1106.
Juola FA, Haussmann MF, Dearborn DC, Vleck CM (2006) Telomere shortening in a long-lived marine bird: cross-sectional analysis and test of an aging tool. *The Auk* **123**, 775-783.
Katona S, Whitehead H (1981) Identifying humpback whales using their natural markings. *Polar Record* **20**, 439-444.
Kent WJ, Sugnet CW, Furey TS*, et al.* (2002) The human genome browser at UCSC. *Genome Research* **12**, 996-1006.
Kim D-H, Doyle MR, Sung S, Amasino RM (2009) Vernalization: winter and the timing of flowering in plants. *Annual Review of Cell and Developmental* **25**, 277-299.
Koch CM, Wagner W (2011) Epigenetic-aging-signature to determine age in different tissues. *Aging (Albany NY)* **3**, 1018-1027.
Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**, 693-705.
Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319**, 1827-1830.
Labbé A (2017) *Effects of age on reproduction and chick rearing in bridled terns (Onychoprion anaethetus) at Penguin Island, Western Australia*, Murdoch University.
Langvatn R, Albon S, Burkey T, Clutton-Brock T (1996) Climate, plant phenology and variation in age of first reproduction in a temperate herbivore. *Journal of Animal Ecology*, 653-670.
Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**, 204-220.
Lee CT (2017) Elasticity of population growth with respect to the intensity of biotic or abiotic driving factors. *Ecology* **98**, 1016-1025.
Leung A, Schones DE, Natarajan R (2012) Using epigenetic mechanisms to understand the impact of common disease causing alleles. *Current opinion in immunology* **24**, 558-563.
Levine ME (2013) Modeling the rate of senescence: can estimated biological age predict mortality more accurately than chronological age? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **68**, 667-674.
Lewison R, Oro D, Godley B*, et al.* (2012) Research priorities for seabirds: improving conservation and management in the 21st century. *Endangered Species Research* **17**, 93-121.
Li L-C, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427-1431.
Li M, Stoneking M (2012) A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biol* **13**, R34.
Li Q, Li N, Hu X*, et al.* (2011) Genome-wide mapping of DNA methylation in chicken. *PLoS One* **6**, e19428.
Lill A, Baldwin J (1983) Weight Changes and the Mode of Depot Fat Accumulation in Migratory Short Tailed Shearwaters. *Australian journal of zoology* **31**, 891-902.
Lipinski M (1986) Methods for the validation of squid age from statoliths. *Journal of the Marine Biological Association of the United Kingdom* **66**, 505-526.
Lister R, Pelizzola M, Dowen RH*, et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322.
Liu C, Marioni R, Hedman ÅK*, et al.* (2016) A DNA methylation biomarker of alcohol consumption. *Molecular psychiatry*.

Liu L, Li Y, Li S*, et al.* (2012) Comparison of next-generation sequencing systems. *BioMed Research International* **2012**.
Liu Q (2017) TMBIM-mediated Ca2+ homeostasis and cell death. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1864**, 850-857.
Lockett GA, Almond EJ, Huggins TJ, Parker JD, Bourke AF (2016) Gene expression differences in relation to age and social environment in queen and worker bumble bees. *Experimental Gerontology*.
Lopatina N, Haskell JF, Andrews LG*, et al.* (2002) Differential maintenance and de novo methylating activity by three DNA methyltransferases in aging and immortalized fibroblasts. *Journal of cellular biochemistry* **84**, 324-334.
Lopez-Serra P, Esteller M (2012) DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene* **31**, 1609-1622.
Louzao M, Bécares J, Rodríguez B*, et al.* (2009) Combining vessel-based surveys and tracking data to identify key marine areas for seabirds. *Marine Ecology Progress Series* **391**, 183-197.
Lowe R, Barton C, Jenkins CA*, et al.* (2018) Ageing-associated DNA methylation dynamics are a molecular readout of lifespan variation among mammalian species. *Genome Biology* **19**, 22.
Luo J, Yu Y, Zhang H*, et al.* (2011) Down-regulation of promoter methylation level of CD4 gene after MDV infection in MD-susceptible chicken line **5**, S7.
Maegawa S, Gough SM, Watanabe-Okochi N*, et al.* (2014) Age-related epigenetic drift in the pathogenesis of MDS and AML. *Genome Research* **24**, 580-591.
Maegawa S, Hinkal G, Kim HS*, et al.* (2010) Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Research* **20**, 332-340.
Maegawa S, Lu Y, Tahara T*, et al.* (2017) Caloric restriction delays age-related methylation drift. *Nature Communications* **8**, 539.
Mallory ML, Gilchrist HG, Janssen M*, et al.* (2018) Financial costs of conducting science in the Arctic: examples from seabird research. *Arctic Science*.
Mallory ML, Robinson SA, Hebert CE, Forbes MR (2010) Seabirds as indicators of aquatic ecosystem conditions: a case for gathering multiple proxies of seabird health. *Marine Pollution Bulletin* **60**, 7-12.
Malone CS, Miner MD, Doerr JR*, et al.* (2001) CmC (A/T) GG DNA methylation in mature B cell lymphoma gene silencing. *Proceedings of the National Academy of Sciences* **98**, 10404-10409.
Marshall AJ, Serventy DL (1956) The breeding cycle of the short-tailed shearwater, *puffinus tenuirostris* (Temminck), in relation to trans-equatorial migration and its environment. **127**, 489-510.
Mas S, Crescenti A, Gassó P, Vidal-Taboada JM, Lafuente A (2007) DNA cards: determinants of DNA yield and quality in collecting genetic samples for pharmacogenetic studies. *Basic & clinical pharmacology & toxicology* **101**, 132-137.
Massot M, Clobert J, Montes-Poloni L*, et al.* (2011) An integrative study of ageing in a wild population of common lizards. *Functional Ecology* **25**, 848-858.
Matsumoto Y, Buemio A, Chu R, Vafaee M, Crews D (2013) Epigenetic control of gonadal aromatase (cyp19a1) in temperature-dependent sex determination of red-eared slider turtles. *PLoS One* **8**, e63599.
Matsumoto Y, Hannigan B, Crews D (2016) Temperature Shift Alters DNA Methylation and Histone Modification Patterns in Gonadal Aromatase (cyp19a1) Gene in Species with Temperature-Dependent Sex Determination. *PLoS One* **11**, e0167362.
Mawlood SK, Dennany L, Watson N, Dempster J, Pickard BS (2016) Quantification of global mitochondrial DNA methylation levels and inverse correlation with age at two CpG sites. *Aging (Albany NY)* **8**, 636.
McDonald PG, Griffith SC (2011) To pluck or not to pluck: the hidden ethical and scientific costs of relying on feathers as a primary source of DNA. *Journal of Avian Biology* **42**, 197-203.

McInnes JC, Alderman R, Deagle BE*, et al.* (2017) Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. *Methods in Ecology and Evolution* **8**, 192-202.
Meissner A, Gnirke A, Bell GW*, et al.* (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Research* **33**, 5868-5877.
Mills MS, Ryan PG (2005) Modelling impacts of long-line fishing: what are the effects of pair-bond disruption and sex-biased mortality on albatross fecundity? *Animal Conservation* **8**, 359-367.
Mohn F, Schübeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends in Genetics* **25**, 129-136.
Møller A, De Lope F (1999) Senescence in a short-lived migratory bird: age-dependent morphology, migration, reproduction and parasitism. *Journal of Animal Ecology* **68**, 163-171.
Musick JA (1999) Ecology and conservation of long-lived marine animals **23**, 1-10.
Nadal J, Ponz C, Margalida A (2018) Population age structure as an indicator for assessing the quality of breeding areas of Common quail (Coturnix coturnix). *Ecological indicators* **93**, 1136-1142.
Nätt D, Agnvall B, Jensen P (2014) Large sex differences in chicken behavior and brain gene expression coincide with few differences in promoter DNA-methylation. *PLoS One* **9**, e96376.
Nelson ME (2002) The science, ethics, and philosophy of tooth extractions from live-captured white-tailed deer: a response to Festa-Bianchet et al.(2002). *Wildlife Society Bulletin (1973-2006)* **30**, 284-288.
Nilsen FM, Parrott BB, Bowden JA*, et al.* (2016) Global DNA methylation loss associated with mercury contamination and aging in the American alligator (Alligator mississippiensis). *Science of The Total Environment* **545**, 389-397.
Nilsson E, Jansson PA, Perfilyev A*, et al.* (2014) Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes* **63**, 2962-2976.
Nussey D, Coulson T, Festa-Bianchet M, Gaillard JM (2008) Measuring senescence in wild animal populations: towards a longitudinal approach. *Functional Ecology* **22**, 393-406.
Nussey DH, Froy H, Lemaitre J-F, Gaillard J-M, Austad SN (2013) Senescence in natural populations of animals: widespread evidence and its implications for bio-gerontology. *Ageing research reviews* **12**, 214-225.
O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. *Molecular Ecology*.
Oliveira DP, Marioni B, Farias IP, Hrbek T (2014) Genetic Evidence for Polygamy as a Mating Strategy in Caiman crocodilus. *Journal of Heredity*, esu020.
Ooi SK, Qiu C, Bernstein E*, et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-717.
Organ CL, Shedlock AM, Meade A, Pagel M, Edwards SV (2007) Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180-184.
Oro D, Hernández N, Jover L, Genovart M (2014) From recruitment to senescence: food shapes the age-dependent pattern of breeding performance in a long-lived bird. *Ecology* **95**, 446-457.
Ou X, Gao J, Wang H*, et al.* (2012) Predicting human age with bloodstains by sjTREC quantification. *PLoS One* **7**, e42412.
Ozgul A, Childs DZ, Oli MK*, et al.* (2010) Coupled dynamics of body mass and population growth in response to environmental change. *Nature* **466**, 482-485.
Pai AA, Bell JT, Marioni JC, Pritchard JK, Gilad Y (2011) A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS Genet* **7**, e1001316.
Park JK, Ryu JK, Lee KH*, et al.* (2007) Quantitative analysis of NPTX2 hypermethylation is a promising molecular diagnostic marker for pancreatic cancer. *Pancreas* **35**, e9-e15.
Parle-Mcdermott A, Harrison A (2011) DNA methylation: a timeline of methods and applications. *Frontiers in Genetics* **2**, 74.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, 637-669.
Parrott BB, Bowden JA, Kohno S*, et al.* (2014) Influence of tissue, age, and environmental quality on DNA methylation in Alligator mississippiensis. *Reproduction* **147**, 503-513.
Patil V, Ward RL, Hesson LB (2014) The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* **9**, 823-828.
Pauli JN, Whiteman JP, Marcot BG, McClean TM, Ben-David M (2011) DNA-based approach to aging martens (Martes americana and M. caurina). *Journal of Mammalogy* **92**, 500-510.
Payo-Payo A, Sanz-Aguilar A, Genovart M*, et al.* (2018) Predator arrival elicits differential dispersal, change in age structure and reproductive performance in a prey population. *Scientific Reports* **8**, 1971.
Pearce-Higgins JW, Eglington SM, Martay B, Chamberlain DE (2015) Drivers of climate change impacts on bird communities. *Journal of Animal Ecology* **84**, 943-954.
Penhallurick J, Wink M (2004) Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome *b* gene. *Emu* **104**, 125-147.
Pérez-Barbería F, Duff E, Brewer M, Guinness F (2014) Evaluation of methods to age Scottish red deer: the balance between accuracy and practicality. *Journal of Zoology* **294**, 180-189.
Petkovich DA, Podolskiy DI, Lobanov AV*, et al.* (2017) Using DNA methylation profiling to evaluate biological age and longevity interventions. *Cell metabolism* **25**, 954-960. e956.
Petzold A, Reichwald K, Groth M*, et al.* (2013) The transcript catalogue of the short-lived fish Nothobranchius furzeri provides insights into age-dependent changes of mRNA levels. *BMC genomics* **14**, 185.
Piatt J, Sydeman W, Sydeman W, Piatt J, Browman H (2007) Seabirds as indicators of marine ecosystems. *Marine Ecology Progress Series* **352**, 199.
Polanowski AM, Robbins J, Chandler D, Jarman SN (2014) Epigenetic estimation of age in humpback whales. *Molecular Ecology Resources* **14**, 976-987.
Popp C, Dean W, Feng S*, et al.* (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101-1105.
Portela A, Esteller M (2010) Epigenetic modifications and human disease. *Nature Biotechnology* **28**, 1057.
Potok ME, Nix DA, Parnell TJ, Cairns BR (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**, 759-772.
Pruitt KD, Tatusova T, Brown GR, Maglott DR (2012) NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Research* **40**, D130-D135.
Prum RO (1999) Development and evolutionary origin of feathers. *The Journal of experimental zoology* **285**, 291-306.
Quach A, Levine ME, Tanaka T*, et al.* (2017) Epigenetic clock analysis of diet, exercise, education, and lifestyle factors. *Aging (Albany NY)* **9**, 419.
Rakyan VK, Down TA, Maslau S*, et al.* (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Research* **20**, 434-439.
Rand AC, Jain M, Eizenga JM*, et al.* (2017) Mapping DNA methylation with high-throughput nanopore sequencing. *Nature Methods*.
Rando TA, Chang HY (2012) Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. *Cell* **148**, 46-57.
Rattiste K, Klandorf H, Urvik J*, et al.* (2015) Skin pentosidine and telomere length do not covary with age in a long-lived seabird. *Biogerontology* **16**, 435-441.
Rawlence NJ, Wood JR, Armstrong KN, Cooper A (2009) DNA content and distribution in ancient feathers and potential to reconstruct the plumage of extinct avian taxa. *Proceedings of the Royal Society of London B: Biological Sciences* **276**, 3395-3402.

Riggs AD (1975) X inactivation, differentiation, and DNA methylation. *Cytogenetic and Genome Research* **14**, 9-25.
Robasky K, Lewis NE, Church GM (2014) The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics* **15**, 56.
Romano A, De Giorgio B, Parolini M*, et al.* (2017) Methylation of the circadian Clock gene in the offspring of a free-living passerine bird increases with maternal and individual exposure to PM 10. *Environmental Pollution* **220**, 29-37.
Rönn T, Volkov P, Davegårdh C*, et al.* (2013) A six months exercise intervention influences the genome-wide DNA methylation pattern in human adipose tissue. *PLoS Genet* **9**, e1003572.
Rosenfeld CS (2010) Animal models to study environmental epigenetics. *Biology of reproduction* **82**, 473-488.
Rossi DJ, Bryder D, Zahn JM*, et al.* (2005) Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proceedings of the National Academy of Sciences* **102**, 9194-9199.
Sarkar G, Turner RT, Bolander ME (1993) Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers. *Genome Research* **2**, 318-322.
Schield DR, Walsh MR, Card DC*, et al.* (2016) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods in Ecology and Evolution* **7**, 60-69.
Scott ME (1988) The impact of infection and disease on animal populations: implications for conservation biology. *Conservation Biology* **2**, 40-56.
Segelbacher G (2002) Noninvasive genetic analysis in birds: testing reliability of feather samples. *Molecular Ecology Notes* **2**, 367-369.
Selman C, Blount JD, Nussey DH, Speakman JR (2012) Oxidative damage, ageing, and life-history evolution: where now? *Trends in Ecology & Evolution* **27**, 570-577.
Sempowski GD, Gooding ME, Liao H, Le PT, Haynes BF (2002) T cell receptor excision circle assessment of thymopoiesis in aging mice. *Molecular immunology* **38**, 841-848.
Seneviratne SS, Jones IL (2008) Mechanosensory function for facial ornamentation in the whiskered auklet, a crevice-dwelling seabird. *Behavioral Ecology* **19**, 784-790.
Shepard EL, Lambertucci SA, Vallmitjana D, Wilson RP (2011) Energy beyond food: foraging theory informs time spent in thermals by a large soaring bird. *PLoS One* **6**, e27375.
Sherley RB, Abadi F, Ludynia K*, et al.* (2014) Age-specific survival and movement among major African Penguin Spheniscus demersus colonies. *Ibis* **156**, 716-728.
Shimoda N, Izawa T, Yoshizawa A*, et al.* (2014) Decrease in cytosine methylation at CpG island shores and increase in DNA fragmentation during zebrafish aging. *Age* **36**, 103-115.
Sih A, Bolnick DI, Luttbeg B*, et al.* (2010) Predator–prey naïveté, antipredator behavior, and the ecology of predator invasions. *Oikos* **119**, 610-621.
Simpson JT, Workman RE, Zuzarte P*, et al.* (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*.
Skira IJ, Brothers NP, Pemberton D (1996) Distribution, abundance and conservation status of Short-tailed Shearwaters Puffinus tenuirostris in Tasmania, Australia. *Marine Ornithology* **24**, 1-14.
Slatko BE, Gardner AF, Ausubel FM (2018) Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology* **122**, e59.
Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT (2018) Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. *Epigenetics & Chromatin* **11**, 25.
Slieker RC, van Iterson M, Luijk R*, et al.* (2016) Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biology* **17**, 191.
Smith L, Burgoyne LA (2004) Collecting, archiving and processing DNA from wildlife samples using FTA® databasing paper. *BMC ecology* **4**, 4.
Smith TB, Marra PP, Webster MS*, et al.* (2003) A call for feather sampling. *The Auk* **120**, 218-221.

Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nature Reviews Genetics* **14**, 204-220.
Soulsbury CD, Lipponen A, Wood K*, et al.* (2018) Age-and quality-dependent DNA methylation correlate with melanin-based coloration in a wild bird. *Ecology and Evolution*.
Souvorov A, Kapustin Y, Kiryutin B*, et al.* (2010) Gnomon–NCBI eukaryotic gene prediction tool. *National Center for Biotechnology Information*, 1-24.
Speller CF, Nicholas GP, Yang DY (2011) Feather barbs as a good source of mtDNA for bird species identification in forensic wildlife investigations. *Investigative genetics* **2**, 1.
Spiers H, Hannon E, Wells S*, et al.* (2016) Age-associated changes in DNA methylation across multiple tissues in an inbred mouse model. *Mechanisms of Ageing and Development* **154**, 20-23.
Steegenga WT, Boekschoten MV, Lute C*, et al.* (2014) Genome-wide age-related changes in DNA methylation and gene expression in human PBMCs. *Age* **36**, 9648.
Stubbs TM, Bonder MJ, Stark A-K*, et al.* (2017) Multi-tissue DNA methylation age predictor in mouse. *Genome Biology* **18**, 68.
Sun Y, Gao Y, Qiao S*, et al.* (2014) Epigenetic DNA methylation in the promoters of Peroxisome Proliferator-Activated Receptor [gamma] in chicken lines divergently selected for fatness1. *Journal of Animal Science* **92**, 48.
Sziráki A, Tyshkovskiy A, Gladyshev VN (2018) Global remodeling of the mouse DNA methylome during aging and in response to calorie restriction. *Aging Cell* **17**, e12738.
Thompson MJ, vonHoldt B, Horvath S, Pellegrini M (2017) An epigenetic aging clock for dogs and wolves. *Aging (Albany NY)* **9**, 1055.
Tissenbaum HA (2012) Genetics, life span, health span, and the aging process in Caenorhabditis elegans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **67**, 503-510.
Tkadlec E, Zejda J (1998) Small rodent population fluctuations: the effects of age structure and seasonality. *Evolutionary ecology* **12**, 191-210.
Trenkel V, Vaz S, Albouy C*, et al.* (2019) We can reduce the impact of scientific trawling on marine ecosystems. *Marine Ecology Progress Series* **609**, 277-282.
Tricola GM, Simons MJ, Atema E*, et al.* (2018) The rate of telomere loss is related to maximum lifespan in birds. *Phil. Trans. R. Soc. B* **373**, 20160445.
Uhlmann S (2003) *Fisheries bycatch mortalities of sooty shearwaters (Puffinus griseus) and short-tailed shearwaters (P. tenuirostris)* Department of Conservation Wellington,, New Zealand.
Untergasser A, Cutcutache I, Koressaar T*, et al.* (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research* **40**, e115-e115.
Valapour M, Guo J, Schroeder JT*, et al.* (2002) Histone deacetylation inhibits IL4 gene expression in T cells. *Journal of allergy and clinical immunology* **109**, 238-245.
Valdes AM, Glass D, Spector TD (2013) Omics technologies and the study of human ageing. *Nature Reviews Genetics* **14**, 601-607.
Varley KE, Gertz J, Bowling KM*, et al.* (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Research* **23**, 555-567.
Varriale A, Bernardi G (2006) DNA methylation in reptiles. *Gene* **385**, 122-127.
Velarde E, Ezcurra E (2018) Are seabirds' life history traits maladaptive under present oceanographic variability? The case of Heermann's Gull (*Larus heermanni*). *The Condor* **120**, 388-401.
Verhulst EC, Mateman AC, Zwier MV*, et al.* (2016) Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Molecular Ecology*.
Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP (2017) DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. *Forensic Science International: Genetics* **28**, 225-236.
Vidal-Bralo L, Lopez-Golan Y, Gonzalez A (2016) Simplified Assay for Epigenetic Age Estimation in Whole Blood of Adults. *Frontiers in Genetics* **7**.

Visser GH (2001) Chick growth and development in seabirds. *Biology of marine birds. Edited by EA Schreiber and J. Burger. CRC Press, Boca Raton, Fla*, 439-465.
Wagner W (2017) Epigenetic aging clocks in mice and men. *Genome Biology* **18**, 107.
Wallis JW, Aerts J, Groenen MA*, et al.* (2004) A physical map of the chicken genome. *Nature* **432**, 761-764.
Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506-513.
Wang M-H, Marinotti O, Zhong D*, et al.* (2013) Gene expression-based biomarkers for Anopheles gambiae age grading.
Warren WC, Clayton DF, Ellegren H*, et al.* (2010) The genome of a songbird. *Nature* **464**, 757-762.
Watanabe YY, Ito M, Takahashi A (2014) Testing optimal foraging theory in a penguin–krill system. *Proceedings of the Royal Society of London B: Biological Sciences* **281**, 20132376.
Waters MF, Minassian NA, Stevanin G*, et al.* (2006) Mutations in voltage-gated potassium channel KCNC3 cause degenerative and developmental central nervous system phenotypes. *Nature genetics* **38**, 447-451.
Weber KL, Bolander ME, Sarkab G (1998) Rapid acquisition of unknown DNA sequence adjacent to a known segment by multiplex restriction site PCR. *Biotechniques* **25**, 415-419.
Weidner CI, Lin Q, Koch CM*, et al.* (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biology* **15**, R24.
Weimerskirch H (2001) Seabird demography and its relationship with the marine environment. In: *Biology of marine birds*, pp. 128-149. CRC press.
Weimerskirch H, Lequette B, Jouventin P (1989) Development and maturation of plumage in the wandering albatross (*Diomedea exulans*). *Journal of Zoology* **219**, 411-421.
Wellbrock AH, Bauch C, Rozman J, Witte K (2012) Buccal swabs as a reliable source of DNA for sexing young and adult Common Swifts (Apus apus). *Journal of Ornithology* **153**, 991-994.
Weng JT-Y, Wu LS-H, Lee C-S, Hsu PW-C, Cheng AT (2015) Integrative epigenetic profiling analysis identifies DNA methylation changes associated with chronic alcohol consumption. *Computers in biology and medicine* **64**, 299-306.
Wetterstrand K (2013) DNA sequencing costs: data from the NHGRI Genome Sequencing Program (GSP). *National Human Genome Research Institute*.
Wilmers CC, Nickel B, Bryce CM*, et al.* (2015) The golden age of bio-logging: how animal-borne sensors are advancing the frontiers of ecology. *Ecology* **96**, 1741-1753.
Wilson DS, Tracy CR, Tracy CR (2003) Estimating age of turtles from growth rings: a critical evaluation of the technique. *Herpetologica* **59**, 178-194.
Wolff GL, Kodell RL, Moore SR, Cooney CA (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *The FASEB Journal* **12**, 949-957.
Wooller R, Bradley J, Croxall JP (1992) Long-term population studies of seabirds. *Trends in Ecology & Evolution* **7**, 111-114.
Wooller R, Bradley J, Skira I, Serventy D (1990) Reproductive success of short-tailed shearwaters (*Puffinus tenuirostris*) in relation to their age and breeding experience. *The Journal of Animal Ecology*, 161-170.
Wright PG, Mathews F, Schofield H*, et al.* (2018) Application of a novel molecular method to age free-living wild Bechstein's bats. *Molecular Ecology Resources*.
Xie W, Barr CL, Kim A*, et al.* (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148**, 816-831.
Xu Q, Zhang Y, Sun D, Wang Y, Yu Y (2007) Analysis on DNA methylation of various tissues in chicken. *Animal biotechnology* **18**, 231-241.
Yan X-p, Liu H-h, Liu J-y*, et al.* (2015) Evidence in duck for supporting alteration of incubation temperature may have influence on methylation of genomic DNA. *Poultry science* **94**, 2537-2545.

Yin Y, Morgunova E, Jolma A*, et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239.
You JS, Kelly TK, De Carvalho DD*, et al.* (2011) OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proceedings of the National Academy of Sciences* **108**, 14497-14502.
Young RC, Kitaysky AS, Haussmann MF*, et al.* (2013) Age, sex, and telomere dynamics in a long-lived seabird with male-biased parental care. *PLoS One* **8**, e74931.
Youngblood B, Hale JS, Ahmed R (2013) T-cell memory differentiation: insights from transcriptional signatures and epigenetics. *Immunology* **139**, 277-284.
Zaghlool SB, Al-Shafai M, Al Muftah WA*, et al.* (2015) Association of DNA methylation with age, gender, and smoking in an Arab population. *Clinical epigenetics* **7**, 1.
Zbieć-Piekarska R, Spólnicka M, Kupiec T*, et al.* (2015a) Examination of DNA methylation status of the ELOVL2 marker may be useful for human age prediction in forensic science. *Forensic Science International: Genetics* **14**, 161-167.
Zbieć-Piekarska R, Spólnicka M, Kupiec T*, et al.* (2015b) Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Science International: Genetics*.
Zhang G, Li C, Li Q*, et al.* (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311-1320.
Zhang M, Yan F-B, Li F*, et al.* (2017) Genome-wide DNA methylation profiles reveal novel candidate genes associated with meat quality at different age stages in hens. *Scientific Reports* **7**.
Zykovich A, Hubbard A, Flynn JM*, et al.* (2014) Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell* **13**, 360-366.

-This page is intentionally blank-